

**HUMAN CHROMOSOME TAGGING IN SOMATIC  
CELLS USING SINE AND LINE TARGETING  
VECTORS**

J. E. Vivienne Watson

PhD.  
University Of Edinburgh

MRC Human Genetics Unit  
Edinburgh  
1994



## **DECLARATION**

I declare that:

- a) this thesis has been composed by myself, and
- b) that the work contained within is my own unless otherwise stated.



## Acknowledgements

This work was funded by an MRC studentship.

First of all I have to thank David Porteous for keeping a bench space for me, always believing that I would eventually arrive. I have appreciated his support and encouragement throughout the project and in particular over the last few months for enthusing me when I was flagging, thank you.

There are several postdocs and others in the Human Genetics Unit who have helped me in various ways; Tony for initiating the project and for both criticism and support of my work; Chris for advice on bugs (both in culture and on the computer) and for stimulating discussions, intellectual and otherwise; big Paul for advice on vector construction (whether it be with Bluescribe or with IslandDraw), for his music machine and for many pints of IPA; Yoshiro for invaluable advice on FISH and Diane for practical support; John Maule for being an oracle on PFGE and advising on meganuclease; Julia D. for her knowledge of gene targeting and for reading a difficult chapter; David K. for encouragement and advice; the office girls Kathy E., Wendy K. and Julia B. for sympathetic encouragement; Sheila C. for ordering and fixing most things in the lab and all the rest of the people in the 'West Wing' for making it a lively and friendly place to work. Thanks also to Professors Andrew Wyllie and Nick Hastie for encouraging words and discussion.

Big thanks are also due to Sandy Bruce and Norman Davidson for reproducing my photographs and autorads included here, and to Sheila Mould for such an efficient library service.

Maria Stack analysed the transfected chromosome 17 hybrids provided by John Boyle (CRC, Manchester), and Peter Teague (MRC, HGU) helped with statistical analysis.

I would especially like to thank my parents for continued support and encouragement throughout my life as a student; Rupert for always believing in me; and Jane 'wee snifter' McNair for being a great person to live with. And last but not at all least David T. for his patience, understanding, and support when it really mattered, and above all for his irrepressible sense of humour.

## Abstract

Site-specific homologous recombination is now an established method for introducing selectable markers into the mammalian genome, but the need to design specific targeting vectors for each locus is a practical limitation. More generalised approaches to chromosome tagging would be valuable.

Preliminary studies from the group suggested that dominant selectable markers are preferentially inserted into human DNA sequences in a somatic cell hybrid when they are flanked by certain human repetitive elements within a vector construct.

To investigate this further I have therefore constructed a series of targeting vectors containing two types of human-specific repeat sequences: Blur8, a short interspersed nuclear element (SINE) from the Alu family of sequences; and two fragments from the 3' end of a long interspersed nuclear element (LINE) L1. The vectors contain dominant markers selectable in mammalian cell culture (*neoR*), as well as in yeast (*Leu2*) and bacteria (*Amp*), and restriction sites to enable unique cleavage of genomic DNA after integration of the vector (*I-SceI* and *lacO*). I have introduced the vectors into J1c4, a somatic cell hybrid containing a single human chromosome in a background of Chinese hamster ovary (CHO) DNA. I have tested the ability of the chosen repeat sequences to mediate integration of the vector into human DNA in preference to Chinese hamster DNA. The sites of integration were localised by fluorescent *in situ* hybridisation (FISH).

Forty-three G418-resistant clones were isolated after targeting with a vector containing two Alu-repeat sequences. Of 37 of these analysed by FISH, only in 1 had the vector integrated into the human chromosome 11. Using vectors containing two L1 fragments or an L1 and an Alu sequence a total of 65 G418-resistant clones were isolated in three different experiments. FISH analysis showed that in 8/14 of these clones the marker was localised to the human chromosome 11. This suggests that fragment(s) from the L1 repetitive element can direct integration of a vector into human DNA within a somatic cell hybrid.

Characterisation of clones containing markers in the human chromosome has been carried out using PCR and restriction endonuclease mapping strategies. Cleavage at a unique restriction endonuclease site introduced by the vector has been demonstrated.

These studies suggest that this strategy has potential for increasing the usefulness of somatic cell hybrids in the analysis of the human genome by enabling the creation of a series of hybrids in which the single human chromosome has been tagged at different locations. Such hybrids could be used directly in mapping around the

tagged region, or they could be used to generate reduced hybrids enriched for a chromosomal region of interest which could then be used in mapping strategies or as cloning resource. It would also facilitate the identification of genes by functional analysis following transfer of selected fragments into an appropriate cellular assay system.

## Abbreviations

$\beta$ -gal	$\beta$ -galactosidase
A	adenine
AMP	adenine monophosphate
<i>Amp</i>	ampicillin
APC	adenomatous polyposis coli
APRT	AMP phosphoribosyl transferase
AT	ataxia telangiectasia
ATP	adenine triphosphate
BCL	Boehringer Mannheim
Bio	biotin
bp	base pairs
BSA	bovine serum albumen
C	cytosine
ccc	covalently closed circular
CCD	charge-coupled device
cDNA	coding DNA
CHO	Chinese hamster ovary
Ci	Curies
CIP	calf intestinal phosphatase
CMGT	chromosome-mediated gene transfer
DAPI	4, 6-diamidino-2-phenylindole
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
dNTP	deoxynucleotide triphosphate
dTTP	deoxythymidine triphosphate
dUTP	deoxyuridine triphosphate
DHFR	dihydrofolate reductase
DIG	digoxigenin
DMEM	Dulbecco's modified minimal essential medium
DMSO	dimethylsulphoxide
DNA	deoxyribose nucleic acid
ds DNA	double stranded DNA
DTT	dithiothreitol
EC	extra chromosomal
<i>Ecogpt</i>	<i>E.coli</i> xanthine-guanine phosphoribosyl transferase
ECR	extra chromosomal recombination
EDTA	ethylenediaminetetra-acetic acid
ERCC	excision repair cross-complementing
ES cell	embryonic stem cell
FCS	foetal calf serum
FISH	fluorescent <i>in situ</i> hybridisation
FITC	fluorescein isothiocyanate
G	guanine
g	grams

G band	Giemsa band
G418	gentamycin 418
G6PD	glucose 6 phosphate
HPRT	hypoxanthine-guanidine phosphoribosyl transferase
HR	homologous recombination
hr	hours
HSV	Herpes-Simplex virus
IFGT	irradiation fusion gene transfer
IGF2	insulin-like growth factor
INS	insulin
IPTG	isopropyl $\beta$ -D-thiogalactopyranoside
IR	illegitimate recombination
J1c4	J1 clone 4
kb	kilobases
L-broth	Luria-broth
L1	Line1
lacO	lac operator
Leu2	leucine
LDS	lithium dodecyl sulphate
LINE	long interspersed element
LOH	loss of heterozygosity
Mb	megabases
min	minutes
ml	millilitres
MMCT	microcell-mediated chromosome transfer
mRNA	messenger RNA
NEB	New England Biolabs
<i>neo</i>	neomycin
<i>neoR</i>	neomycin-resistance gene
ng	nanogram
°C	degrees Celcius
OD	optical density
oligo	oligonucleotide
ORF	open reading frame
PBS	phosphate-buffered saline
pBS	Bluescribe plasmid
PCR	polymerase chain reaction
PEG	polyethylene glycol
PI	propidium iodide
PNS	positive-negative selection
Pol III	RNA polymerase III
R band	reverse band
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RNase	ribonuclease
RNP	ribonuclear protein
rpm	revolutions per minute
s	seconds
SDS	sodium dodecyl sulphate

SINE	short interspersed nuclear element
ss DNA	single-stranded DNA
STS	short tagged sequence
T	thymine
TK	thymidine kinase (human)
tk	thymidine kinase (HSV)
Topo I	topoisomerase I
tRNA	transfer RNA
UV	ultra violet
X-gal	5-bromo-4-chloro-3-indolyl $\beta$ -D-galactopyranoside
XP	xeroderma pigmentosum
YAC	yeast artificial chromosome

## List of Figures

### Chapter 1

1.1 The structural relationship of Alu, B1, and 7SL RNA sequences.....	19
1.2 Multiple sequence alignment of 7SL RNA, Alu consensus, and B1 consensus .....	21
1.3 A Line1 repeat element.....	29
1.4 The double stranded break repair (dsbr) model .....	34
1.5 The Meselson-Radding model .....	34
1.6 The single strand (ss) annealing model for recombination.....	36
1.7 Recombination mechanisms for different types of targeting construct.....	41

### Chapter 3

3.1.0 Construction of pBLOB3XN <sup>-</sup> , pBLOB4XN <sup>-</sup> , and pTVL1.3 .....	81
3.2.0 Construction of 1 $\Omega$ 2SstII .....	83
3.2.1 Sequence of p1 $\Omega$ 1 and p1 $\Omega$ 2.....	85
3.3.0 Construction of pTVBlur1 and pTVBlur2 .....	86
3.3.1 Restriction maps of the vectors pTVBlur1 and pTVBlur2.....	88
3.4.0 Construction of pLSst3 .....	89
3.5.0 Construction of pLL and pLB.....	91
3.5.1 Restriction digests of targeting vectors .....	92
3.5.2 Restriction maps of the vectors pLL and pLB .....	93
3.6 Possible mechanisms for vector integration.....	95

### Chapter 4

4.1.1 Bal31 exonuclease digestion of pTVBlur1 and pTVBlur2.....	97
4.2 FISH of clones L1b1, L1b4, L1b12, L1b11, and L1b3.....	106
4.3 FISH of clones 1.11, 1.14, 1.17, and 1.18 (pTVBlur1 transfectants).....	107
4.4 FISH of clones 1.23, 2.37, and 1.22 (pTVBlur1 and 2 transfectants) .....	108
4.5 FISH of clones L1b5, L1b4, L1b8, and L1b2.....	109
4.6 FISH of clones LB1, LL2, LB24, and LB5 .....	110
4.7 FISH of human lymphocyte metaphases and clones LL20 and LB20.....	111
4.8 FISH of clone LB15 .....	112

### Chapter 5

5.1.1 Positive/ negative selection targeting vectors .....	120
5.1.2 Southern blot hybridisation results of BglII-digests of G418- resistant clones.....	121
5.2 Map of linearised vector pTVBlur showing the position of probes X and Y .....	123

5.3 Restriction enzyme analysis of selected clones from transfection with pTVBlur1.....	125
5.4 Map of linearised vectors pLL and pLB showing position of probes X, Y and Z .....	128
5.5 Southern blot analysis of clones from transfections with vectors pLL and pLB.....	129
5.6 NsiI restriction analysis of selected clones.....	131
5.7 Restriction analysis of clones LB1, LB5, LB15, and LB20.....	133
5.8 I-SceI restriction analysis of clones LB5 and LB20.....	135
5.9 Restriction maps of integration sites in clones LB5 and LL2 .....	137
5.10 Restriction analysis of clones LB24, LL2, and LL20 .....	138
5.11 Detection of <i>neoR</i> and <i>Leu2</i> in pLL and pLB targeted clones by PCR .....	140
5.12 Detection of the loci WT1 and D11S533 in selected clones by PCR.....	145
Chapter 6	
6.1 Anchored long-range mapping with unique restriction sites.....	158



**List of Tables**

Chapter 1

1.1 Characteristics of Alu and Line1 human-specific repeat families .....32

Chapter 3

3.1 Oligonucleotides used in vector construction and sequencing .....80

Chapter 4

4.1 Conditions and results for transfection of J1c4 with vectors L1a, L1b and pTBPNS3 .....98

4.2 Conditions and results for transfection of J1c4 with vectors pTVBlur1 and pTVBlur2 .....99

4.3 Conditions and results for transfection of GM10321 and GM10498 with vectors pTVBlur1 and pTVBlur2 .....101

4.4 Conditions and results for transfection of J1c4 with pLL and pLB .....101

Chapter 5

5.1 Summary of restriction fragment sizes and FISH results for clones arising from transfection of J1c4 with pL1b .....122

5.2 Summary of molecular analysis of isolated clones .....142

5.3 Summary of PCR results to identify p or q arm loss from the human chromosome 11 in selected clones.....146

## Contents

Declaration .....	i
Acknowledgements .....	ii
Abstract .....	iii
List of abbreviations.....	v
List of figures .....	viii
List of tables .....	x
Contents .....	xi
Chapter1 Introduction	
1.1 Somatic cell hybrids and gene cloning.....	1
1.1.1 Somatic Cell Hybridisation .....	1
1.1.2 Somatic cell hybrids and gene mapping.....	3
1.1.3 Somatic cell hybrids as a cloning resource .....	10
1.1.4 Phenotypic complementation and gene cloning.....	10
1.1.5 Localisation of tumour-suppressor genes by phenotypic complementation.....	12
1.2 Experimental Rationale.....	17
1.3 Repetitive DNA Sequences in Mammalian DNA.....	18
1.3.1 The origin of SINEs: Alu and B1 families.....	18
1.3.2 The structure of SINEs : Alu and BI elements.....	20
1.3.3 Dispersal of SINEs.....	20
1.3.4 Isolation and characterisation of Alu repeat sequences .....	23
1.3.5 Evolution of Alu Subfamilies .....	25
1.3.6 Inactivation of Alu Sequences .....	26
1.3.7 Is there a function for Alu and B1 repeat sequences?.....	27
1.3.8 LINES .....	28
1.3.9 The Line 1 (L1) repeat family .....	28
1.3.10 L1 repeats are retroposons .....	30
1.3.11 Distribution of L1-H elements .....	30
1.4 SINE and LINE sequences used in the targeting vector.....	31
1.4.1 Blur8 .....	31
1.4.2 L(1.2), L(1.8) and L(1.4).....	31
1.4.3 Repeat sequences and targeting vectors .....	31
1.5 Recombination in Mammalian cells .....	33
1.5.1 Models for homologous recombination .....	33
1.5.2 Extra chromosomal recombination in mammalian cells.....	35

1.5.3 Recombination between an endogenous locus and transfected DNA .....	37
1.5.4 Illegitimate recombination .....	38
1.5.5 Gene targeting in mammalian cells .....	40
1.5.6 Insertion and replacement targeting vectors .....	40
1.5.7 Mechanisms of gene targeting .....	42
1.6 Chromosome targeting vectors .....	43
1.6.1 Vector design .....	43
1.6.2 Alu repeats and recombination .....	43
1.6.3 L1 repeats and recombination .....	46
1.7 Vector features for manipulation of human genomic DNA .....	46
1.7.1 Sites for unique cleavage .....	46
1.7.3 Affinity Purification .....	47
1.7.4 Leu2 .....	48
<b>Chapter 2 Materials and Methods</b>	
2.1 Bacterial Culture .....	49
2.1.1 Media and additives .....	49
2.1.2 Bacterial Strains .....	49
2.1.3 Plasmid vectors .....	50
2.2 Transformation of bacteria .....	50
2.2.1 Preparation of cells for heat shock transformation .....	50
2.2.2 Heat shock transformation of bacteria .....	51
2.2.3 Preparation of competent cells for electro-transformation .....	51
2.2.4 Transformation of bacterial cells by electroporation .....	52
2.3 Preparation of Plasmid DNA .....	52
2.3.1 Small-scale plasmid DNA preparation .....	52
2.3.2 Midi-scale plasmid DNA preparation .....	53
2.3.3 Large scale preparation of plasmid DNA .....	54
2.4 Enzymatic manipulation of plasmid DNA .....	55
2.4.1 Restriction endonuclease digestion of DNA .....	55
2.4.2 Dephosphorylation of linear plasmid DNA .....	55
2.4.4 Annealing of oligonucleotides .....	56
2.4.5 Conversion of 5' overhangs to blunt-ended molecules .....	56
2.4.6 Ligation of sticky-ended molecules .....	56
2.4.7 Ligation of linkers (sticky-ended) .....	57
2.4.8 Ligation of blunt-ended molecules .....	57
2.4.10 Digestion with I-SceI .....	57

2.5 Preparation of vector DNA for transfection.....	58
2.5.1 Restriction digestion .....	58
2.5.2 Bal 31 exonuclease digestion.....	58
2.5.3 Purification and concentration of Bal31-treated DNA.....	58
2.6 Mammalian Cell Culture.....	59
2.6.1 Cell lines .....	59
2.6.2 General Maintenance .....	59
2.6.3 Transformation of cells by lipofection.....	60
2.6.4 Selection conditions .....	60
2.6.5 Picking resistant clones.....	61
2.7 Manipulation of Genomic DNA .....	61
2.7.1 Rapid isolation of DNA from mammalian cells .....	61
2.7.2 Isolation of high quality DNA from mammalian cells.....	61
2.7.3 Restriction endonuclease digestion of genomic DNA .....	62
2.7.4 Preparation of chromosomal DNA in agarose microbeads.....	63
2.7.8 Restriction digestion of microbeads for pulsed field gel electrophoresis .....	64
2.8 Electrophoresis.....	64
2.8.1 Agarose gel electrophoresis .....	64
2.8.2 Preparative agarose gel electrophoresis .....	65
2.8.3 Pulsed Field Gel Electrophoresis .....	65
2.9 Transfer of DNA to membranes.....	66
2.9.1 Southern transfer of DNA .....	66
2.9.2 Colony transfer.....	66
2.10 Radiolabelling of DNA .....	67
2.10.1 Random Priming of DNA Probes .....	67
2.10.2 End-labelling of DNA oligonucleotides .....	67
2.11 Hybridisation of membranes .....	67
2.11.1 Hybridisation solutions .....	67
2.11.2 Pre-hybridisation .....	68
2.11.3 Hybridisation.....	68
2.11.4 Detection of hybridization .....	69
2.11.5 Re-using filters.....	69
2.12 Sequencing of DNA .....	69
2.12.1 Double strand DNA sequencing .....	69
2.12.2 Polyacrylamide Gel Electrophoresis .....	70

2.13 Polymerase Chain Reaction .....	70
2.13.1 Oligonucleotide preparation.....	70
2.13.2 PCR with Genomic DNA.....	71
2.13.3 Conditions and primers for PCR analysis of targeted cell lines.....	71
2.14 Fluorescence in situ hybridisation (FISH) .....	73
2.14.1 Metaphase Spreads from cell cultures .....	73
2.14.2 Preparation of extended chromosomes and a high density of metaphases .....	73
2.14.3 Production of digoxigenin- or biotin-labelled probe by nick-translation.....	74
2.14.4 Check of probe incorporation of label .....	75
2.14.5 Check of probe length .....	75
2.14.6 Direct labelling of probes with fluorochromes .....	75
2.14.7 Hybridisation with pTVLO $\Omega$ , and pHS53 .....	76
2.14.8 Detection with fluorescein isothiocyanate (FITC) .....	76
2.15 Microscopy.....	77
2.15.1. Confocal microscopy .....	77
2.15.2 Multicolour FISH.....	78
Chapter 3 Vector Construction	
3.0 Preface.....	79
3.1 Construction of pBLOB3XN-, pBLOB4XN- and pTVL1.3 .....	79
3.2 Construction of 1 $\Omega$ 2SstII .....	84
3.3 Construction of pTVBlur1 and pTVBlur2 .....	87
3.4 Construction of pLSst3 .....	87
3.5 Construction of pLL and PLB .....	90
3.6 Mechanisms of recombination .....	94
Chapter 4 transfection of constructs and localisation of integration sites	
4.1 Transfection of vector constructs into J1 c4 .....	96
4.1.1 Preparation of vector DNA for transfection.....	96
4.2 Localisation of vector integration sites by fluorescence in situ hybridisation (FISH) .....	102
4.2.1 The sensitivity of FISH .....	102
4.2.2 Alpha satellite DNA as a probe for FISH .....	103
4.2.3 Optimisation of conditions for FISH .....	104
4.2.4 FISH results.....	105

4.3 Summary of results for FISH analysis of targeted clones .....	113
4.3.1 Vector pL1b (1).....	113
4.3.2 Vectors pTVBlur1 and pTVBlur2.....	113
4.3.3 Vector pL1b (2).....	114
4.3.4 Vectors pLL and pLB.....	115
Chapter 5 Molecular analysis of targeted clones	
5.1. Restriction enzyme analysis.....	118
5.1.0 Preface.....	118
5.1.1 Results with the vector L1b .....	119
5.1.2 Results with the vectors pTVBlur1 and pTVBlur2.....	124
5.1.3 Results with the vectors pLL and pLB.....	130
5.2 Analysis of pLL- and pLB-targeted clones by PCR .....	141
5.3 Summary of restriction enzyme analysis and PCR data for LB and LL .....	
clones .....	144
5.4. Identification of fragments lost from the human chromosome in targeted clones .....	146
Chapter 6 Discussion and Conclusions	
6.1 Integration of the vectors: analysis by FISH .....	148
6.1.1 Sites of integration .....	148
6.1.2 Chromosome shortening and stability.....	149
6.2 Integration of the vectors: molecular analysis.....	152
6.3 Line-1 elements and species-specific recombination.....	153
6.4 Future Studies .....	155
6.5 Manipulation of hybrid clones using novel features introduced by the vector.....	157
6.6 Applications for clones containing an exogenous selectable marker.....	159
References .....	161

## **CHAPTER 1**

# 1 Introduction

## 1.1 Somatic cell hybrids and gene cloning

### 1.1.1 Somatic Cell Hybridisation

#### *The first somatic cell hybrids*

The construction of cell lines which contain all or part of the genetic component of two originally distinct and different somatic cell lines has become fundamental in the analysis of the human genome. The first demonstration of somatic cell hybridization was reported in 1960 (Barski *et al.* 1960; English translation 1961). It was observed that after a period of co-culture of two distinct mouse cell lines a new cell type had spontaneously arisen which contained the combined genomes of the two different parental cells. These new hybrid cells contained a total number of chromosomes nearly equal to the sum of the two parents and contained all the marker chromosomes which characterised the parental cells, reviewed by Ephrussi (1972).

The potential of these hybrids for the study of mammalian genes was immediately appreciated, but it was not until 1964 that a system to enrich for hybrid cells above the background of parental cells was described. Littlefield (1964) was the first to describe an appropriate selection system. He used two mouse cell lines which contained different selective markers: one was deficient in hypoxanthine-guanidine phosphoribosyl transferase (HGPRT or HPRT) and the other in thymidine kinase (TK). Both these enzymes are required to enable a cell to synthesise nucleotides from the nucleosides hypoxanthine and thymidine when the *de novo* nucleotide synthetic pathway is blocked by aminopterin. Therefore neither of the parental cell lines used in this experiment could grow on a medium containing hypoxanthine, aminopterin, and thymidine (HAT medium), but in the hybrid cells both wild type genes are present so they are able to grow and are thus selected for. HAT selection for TK or HPRT has proved a valuable enrichment procedure and is still widely in use today. Davidson and Ephrussi (1965) developed a 'half-selective' technique, whereby unmarked diploid cells could be used to create a hybrid with a heteroploid cell line carrying a marker which could be selected against. The hybrid cells were selected for their ability to grow on the selective medium in a far more proliferate fashion than the wild-type diploid line. This meant that somatic cell hybridization techniques could be applied to a wider range of cell lines.

#### *Inter-specific somatic cell hybrids*

In 1965, Harris and Watkins produced the first inter-specific somatic cell hybrid cell lines between mouse and human cells. They applied a technique first described by



Okada *et al.* in 1958, and used Sendai virus (HVJ virus) to promote fusion between the co-cultured somatic cells. Polyethylene glycol (PEG) was later shown also to induce cell fusion and is easier to use (Pontecorvo *et al.* 1977). The ability to create hybrids between somatic cells from different mammalian species opened up a whole new field in the search for mammalian genes and knowledge of their function.

The potential of somatic cells was widely recognised and developed by many groups, especially in the field of human genetics where gene localisation had previously been limited to studies of inheritance patterns between individuals. Between 1911 and 1967, less than one hundred genes had been assigned to the X chromosome, and very few had been assigned to the autosomes. Some genes had been assigned to specific autosomes by their association with an identifiable chromosomal abnormality, but this approach was obviously limited. The development of inter-specific human hybrids meant that the human genome could be sub-divided to allow the characterisation of a few chromosomes at a time, segregated into a rodent background, (reviewed by Kao 1983). By 1985, 831 cloned genes and polymorphic markers had been mapped in the human genome, over 75% of these had been assigned using somatic cell genetics (HGM 8, 1985).

Ephrussi and Weiss (1965) showed that the 'half-selective' system they had pioneered for intra-specific hybrids could be applied in a similar way to inter-specific hybrids, and in 1967 Weiss and Green used a selection technique to produce mouse-human hybrids which had preferentially lost most of the human component, stably retaining only 1-3 human chromosomes which complemented the tk<sup>-</sup> phenotype of the mouse host. Employing a similar strategy, Migeon and Miller (1968) were able to assign the human gene for thymidine kinase to a single specific chromosome, later designated chromosome 17. This was one of the first examples of human gene assignment using somatic cell hybrids. It utilised the recessive mutant phenotype of one cell to localise the wild type gene complementing the defect from the other cell type, correlating the presence of a gene with a particular chromosome.

Kao *et al.* (1968) developed a method of isolating auxotrophic mutant Chinese hamster cell lines which could be used in such complementation mapping experiments. They induced single gene mutations by exposing cells to sub-lethal doses of chemical mutagens, and then enriched for auxotrophic mutants by growing the cells on medium lacking certain nutrients but which contained the thymidine analogue 5-bromodeoxyuridine (BUdR). Those cells which could grow on the minimal medium incorporated BUdR into their DNA and were subsequently destroyed upon exposure to UV-light, leaving the deficient mutants which were rescued by replacing the medium with one enriched with various metabolites. They used the Chinese hamster

ovary cell line CHO-K1 derived from a cell which had undergone spontaneous transformation in culture to create these mutants. CHO-K1 has many properties that make it amenable to genetic experiments including a short generation time and the ability to readily hybridise with human cells, showing a preferential loss of non-essential human chromosomes.

Since these initial experiments, they and others researchers have derived mutant CHO lines deficient in many different biochemical processes whose complementation is selectable *in vitro*. Using these lines in the creation of human intra-specific hybrids, many different human genes involved in these pathways have been localised and subsequently identified (reviewed by Puck and Kao 1982).

The application of ouabain at a specific concentration has subsequently been shown to kill off human fibroblasts but leave rodent/human hybrid cells unaffected and has been used as a method of enrichment for hybrid cells (reviewed by Goss 1978).

#### *Characterisation of somatic cell hybrids*

Improvements to techniques of chromosome identification, for example Giemsa banding (Sumner *et al.* 1971), facilitated cytogenetic analysis of hybrid cells, allowing unequivocal identification of the chromosomes present, and the systematic Denver karyotyping scheme was adopted. Techniques to determine the presence of genes not directly selectable by biochemical means have evolved dramatically since the first hybrids were described. Initially, gene products were identified by isozyme analysis, exploiting the difference in electrophoretic mobility between enzymes catalysing the same biochemical reactions in rodent and human cells. Alternatively, the presence of genes encoding cell surface markers could be identified using specific antisera, as human-specific antigens are displayed on the surfaces of rodent-human hybrid cells (Puck *et al.* 1971, reviewed by Tunnacliffe *et al.* 1988). The advent of Southern blotting (1975) and filter hybridization greatly improved the speed and sensitivity of detection of specific gene sequences or other anonymous human DNA sequences in the hybrid cell genomic DNA. More recently the polymerase chain reaction and fluorescence in situ hybridization have been used for the rapid characterisation of somatic cell hybrids (described below).

#### **1.1.2 Somatic cell hybrids and gene mapping**

Co-segregation of human-specific genes within somatic cell hybrids was recognised as a way of establishing physical linkage between those genes (Nabholz *et al.* 1969). Hybrid cell lines, therefore, allow mapping of genes by synteny, providing 'parasexual' means to establish linkage between human genes in a fashion comparable to that

which can be achieved in other organisms where large numbers of specific matings can be set up (reviewed by Ruddle (1981)). Linkage through human pedigrees is limited by the small number of available meioses and therefore recombination events.

Monochromosome hybrids have proved useful for efficient and systematic gene mapping. Such hybrids have been established by cell fusion techniques. Sometimes the stable maintenance of a single human chromosome is fortuitous and does not involve biochemical selection, for example chromosome 11 is the sole human chromosome retained in the J1 series of CHO-human hybrids, established by Kao *et al.* (1976). However in most cases, a selection for the human chromosome is required for it to be stably retained in the hybrid. This may be either a dominant-acting selectable marker in the human chromosome, or a host cell mutation which is complemented by a human chromosome.

#### *Irradiation fusion gene transfer (IFGT)*

Once a gene has been assigned to a particular chromosome, the next step is to localise it to within a particular region of that chromosome. Spontaneously arising translocated chromosomes have been observed within the karyotypes of many human individuals which may segregate with a disease phenotype due to disruption of a specific gene at the point of translocation. Translocations may also be induced in culture. Spontaneous, and often disease associated, chromosome deletions also occur. Such rearranged chromosomes with or without a selectable marker can be segregated singly into a somatic cell hybrid and stably maintained. Markers can therefore be mapped proximal or distal to a translocation breakpoint, or within or outside of a deleted region, allowing regional assignment of markers and at the same time defining the chromosomal rearrangement.

A limitless number of breaks, translocations and other chromosome rearrangements can be induced by exposing cells to  $\gamma$ -irradiation. In 1975 Goss and Harris reported that subjecting human male lymphocytes with large doses (1-4 krad) of ionising radiation before fusing them with rodent cells created clones containing differing sub-fragments of human chromosomes. By growing the hybrids on HAT medium they were able to select for those clones retaining fragments of the human X chromosome surrounding the gene for HPRT. A map of the relative order of linked genes G6PD, PGK and  $\alpha$ -gal could be established, and their distance from HPRT estimated by the frequency with which each of the markers was found to co-segregate. The number of new linkage groups that could be established using this technique at that time was limited only by the numbers of endogenous markers available to characterise the sub-chromosomal fragments.

The technique of irradiation fusion gene transfer (IFGT) can be applied to hybrids that originally contain a single human chromosome. Irradiation of a monochromosome hybrid followed by fusion with a non-irradiated rodent recipient will give rise to new hybrids containing multiple fragments of human DNA from only the chromosome of interest. This simplifies to a large extent the production of maps from IFGT hybrids. However, IFGT usually produces hybrids containing multiple non-contiguous fragments, some of which are non-selected.

With the advent of powerful techniques for DNA and chromosome analysis such as the polymerase chain reaction (PCR), and fluorescence *in situ* hybridisation (FISH), a rapid characterisation of these hybrids has become possible, as demonstrated by Glaser *et al.* (1990). The density of genes and anonymous DNA sequences already assigned only extends the resolution of maps which can be created using these hybrids, an example of this is the linkage map created of the X chromosome by Benham *et al.* 1989. The group of Cox *et al.* (1990) was one of the first to realise the full potential in the combination of new techniques for rapid analysis with IFGT for the mapping of genes on a large scale. As reviewed by Cox and Myers (1992) and Walter and Goodfellow (1993), Cox *et al.* have adapted the strategy of Goss and Harris (1975) to perform large scale mapping of human genes on irradiation hybrids, using the principle of conservation of synteny for closely linked markers. Using PCR, the presence, or absence, of anonymous marker sequences or 'sequence tagged sites' (STSs) can rapidly be determined for each hybrid. If enough hybrids are produced and analysed in this way, their precise human DNA content is not important; any conflicting data arising from non-contiguous chromosome fragments in a particular hybrid will become statistically insignificant when taken as part of a large scale analysis. For example, Goold *et al.* (1993) from the group of Cox, report that using 86 independent irradiation hybrids of chromosome 4, a total of 437 STSs have been distinguished from each other and 195 of these have been ordered with respect to each other, creating a map of the entire chromosome 4 with a resolution of about 1Mbp.

Taking a slightly different approach, Kao *et al.* (1976) generated a series of hybrids containing deletions in a specific human chromosome. They induced breaks in the DNA of a hybrid containing a single human chromosome 11 using mutagens such as ethylmethanesulphonate (EMS) and acridine mustard. By imposing selection against the human antigen system  $A_L$  using specific antisera, they were able to create a set hybrids with different, but nested deletions for this region. More recently, Glaser *et al.* (1990) used these hybrids, and others similarly generated, as a panel of DNAs to enable high resolution mapping of the region 11p13-11p14.



### *Chromosome-mediated gene transfer*

Isolated human metaphase chromosomes can be directly introduced into rodent cells to create inter-specific somatic cell hybrids. An alternative strategy to irradiation and fusion of cells, this method allows the direct production of hybrids containing sub-chromosomal human DNA fragments. Known as chromosome-mediated gene transfer (CMGT), this technique was pioneered by McBride and Ozer in 1970 and is reviewed by Klobutcher and Ruddle (1981). It was used by Klobutcher and Ruddle (1979) to create hybrids containing fragments of human fibroblast DNA selected for their ability to complement the  $tk^-$  phenotype of the recipient mouse cell. As well as selecting for complementation of an endogenous mutant phenotype such as  $tk^-$  or  $HPRT^-$ , selection can be applied for a specific cell surface antigen using antisera and complement (Pritchard and Goodfellow 1986).

Shih *et al.* (1981) showed that activated oncogenes have dominant transforming capabilities in somatic cell hybrids, as DNA from carcinomas and neuroblastomas was able to induce transformation to tumorigenicity in NIH3T3 cells. It was this property that allowed the isolation of an oncogene from the T24 bladder carcinoma, as described by Pulciani *et al.* (1982) and Cooper *et al.* (1984). Activated oncogenes in the donor DNA can be used as selectable markers during CMGT (Porteous *et al.* 1987, reviewed by Porteous 1987). In these experiments, CMGT was used to transfer chromosome fragments from a human bladder carcinoma cell line (EJ) to recipient mouse cells. The EJ bladder carcinoma cell line contains an activated HRAS1 oncogene which gives it its tumorigenic properties (Shih and Weinberg 1982). By selecting for hybrid cells which showed anchorage-independent growth they were able to isolate a series of clones containing fragments of DNA from around the HRAS1 locus on chromosome 11p. The potential of CMGT as a mapping tool was clearly demonstrated by Xu *et al.* 1988, when they constructed a genetic map of chromosome 17 using hybrids selected for the transfer of the TK gene. However, co-transfer of non-selected fragments is also observed. In most, if not all cases, hybrid cell lines produced by CMGT become karyotypically stable only after the donor DNA has acquired host telomere function i.e. after it has integrated into a host chromosome. The stabilisation process is thought to involve breakage, deletions and other rearrangements in the transforming DNA before its eventual integration; this explains the presence of non-contiguous, non-selected DNA in these hybrids. These hybrids, however, can now be rapidly and extensively characterised using techniques such as PCR, as described for IFGT hybrids, which improves their usefulness in gene mapping. Presence and absence mapping using 'panels' of DNA from these hybrids is still useful, but statistical methods need to be applied for linkage analysis. Caution

should be exercised before using these reduced hybrids for enrichment cloning, (discussed below), as clones containing non-contiguous DNA sequences may be produced.

### *Fluorescence in situ hybridisation (FISH)*

*In situ* hybridisation can be used to complement somatic cell hybrid techniques in mapping. As early as 1981, it was reported that radio-labelled probes of novel gene sequences and anonymous probes can be given a map position directly by virtue of their ability to hybridise to the complementary DNA sequence and be detected on preparations of whole metaphase (or interphase) chromosomes (Gerhard *et al.* 1981). Since then, the use of fluorochromes has superseded that of radioactive isotopes in labelling of probes for *in situ* hybridisation, and, combined with improvements in visualisation systems, has increased the sensitivity and speed of signal detection, making it a widely applicable technique. The incorporation of antigens such as biotin and digoxigenin into a probe enables multiple layers of specific fluorochrome-tagged antibodies to be built up so that the signal from the probe is amplified sufficiently to enable it to be detected under UV light using a microscope (Pinkel *et al.* 1986).

Gene mapping by FISH compares favourably to that using hybrid panels. A probe complementary to a single copy sequence can be localised using a single spread of metaphase chromosomes from normal human lymphocytes, instead of it requiring  $\sim 10^6$  copies of the sequence ( $\sim 1\mu\text{g}$  of hybrid DNA) on a hybrid panel Southern blot. Using FISH, the probe is also immediately assigned a sub-chromosomal location without the need for a characterised set of hybrids or the identification of syntenic markers. A more precise localisation of probes is obtained in the same way as is used on hybrid panels, by mapping them relative to probes of known location. However, for reliable and consistent results, FISH requires cloned sequences of between 30-50 kb as probes (often cosmid clones are used), whereas probe length is less restrictive with somatic cell hybrid mapping. The most useful approach is to integrate both of these methods, combining the reliability and reusability of well characterised hybrid panels with the power of FISH.

Techniques developed to extend the length of chromosomes used for FISH have improved the resolution of probe localisation and preparations of interphase nuclei have been used to estimate distances between probes providing relative mapping resolutions of up to 50 kilobase pairs (Lawrence *et al.* 1988). More recently, FISH to 'free' chromatin fibres in interphase nuclei has permitted probes as little as 21kb apart to be distinguished (Heng *et al.* 1992).

### *Microcell-mediated chromosome transfer*

The method of microcell-mediated chromosome transfer (MMCT) has been developed to introduce either whole chromosomes or sub-chromosomal fragments in an intact form into rodent recipient cells. As described by Fournier and Ruddle (1977), micronuclei, (induced by extended exposure to colcemid), are purified from donor cells and subsequently fused with recipient cells, to produce hybrids containing only a subset of the genetic material of the donor cell. In order for transfected chromosomes to be specifically and stably maintained, the donor DNA must contain a marker gene, selectable in the hybrid cell.

Dowdy *et al.* (1990) have developed a technique combining MMCT and IFGT which they call 'XMMCT' to introduce intact derived chromosomes into hybrid cells. Microcells prepared from monochromosome hybrids are subjected to a dose of  $\gamma$  radiation such that the chromosomes in the microcells are broken in a limited fashion, prior to fusion with the recipient cells. The original monochromosome hybrid used in their experiment contains a naturally occurring translocation chromosome t(Xqter→Xq21: :11q13→11pter) as its sole human component, and contains the gene for HPRT. This means that hybrids containing this chromosome are selectable on HAT medium. Hybrids produced by XMMCT from this monochromosome hybrid were therefore selected on HAT to enrich for those clones which contained t(11:X) chromosome with limited rearrangements. They were characterised cytogenetically and by RFLP analysis using probes from 11p. Dowdy *et al.* identified hybrids with both single and multiple interstitial deletions in 11p. These radiation-reduced der(11) chromosomes have proved useful in functional analyses of the involvement of 11p in tumour suppression (Dowdy *et al.* 1991, L. Reid pers. comm., described below).

### *DNA-mediated gene transfer*

Wigler *et al.* (1977) demonstrated that the thymidine kinase (tk) gene from Herpes Simplex virus (HSV) could be transferred to a mouse tk<sup>-</sup> cell directly through the introduction of restriction endonuclease-cleaved HSV DNA. Transfected cells were selected for restoration of tk function. Subsequent experiments from the same lab (Wigler *et al.* 1979) showed that DNA from a wild type eukaryotic cell could be used to restore the function of a recipient mouse cell, deficient in AMP phosphoribosyl transferase (APRT). Successful gene transfer was achieved with DNA from Chinese hamster, mouse, and human cells. These initial studies of DNA mediated gene transfer (DMGT) were limited to cell lines with endogenous markers.

### *Bacterial-selectable markers in mammalian cells*

In 1980, Mulligan and Berg demonstrated that the gene from *Escherichia coli* encoding xanthine-guanine phosphoribosyl transferase (*Ecogpt*) could be transferred and expressed in mammalian cells to an extent that it could complement the HPRT<sup>-</sup> phenotype of those cells. Srivatsan *et al.* (1984) showed that a plasmid containing *Ecogpt* can stably integrate at random into human chromosomes.

These advances meant that markers could be introduced into previously non-selectable regions of the human genome, allowing them now to be enriched in somatic cell hybrids through methods of chromosome transfer. This was put into practice by Saxon *et al.* (1985) who describe a strategy for creating monochromosome hybrids containing any human chromosome by tagging them with an exogenous introduced marker (reviewed by Saxon and Stanbridge 1987). Human HPRT<sup>-</sup> cells were transfected with a plasmid carrying *Ecogpt*. Single, intact chromosomes were introduced into HPRT<sup>-</sup> rodent cells from transfected HPRT<sup>+</sup> human cells by MMCT. Only those chromosomes carrying a stably-integrated, functional *Ecogpt* gene would form a hybrid which could survive on HAT medium. Conditions could be manipulated so that the new hybrids contained a single human chromosome. Their protocol was, however, limited to transfer of chromosomes from those strains which are initially HPRT<sup>-</sup>. Mycophenolic acid (MPA) can inhibit the action of HPRT in mammalian cells, therefore, selection for *Ecogpt* can be obtained in HPRT<sup>+</sup> cells using HAT medium supplemented with MPA (Mulligan and Berg 1981).

An alternative, and now very widely used, bacterial marker gene shown to function in mammalian cells is the *E.coli* gene for aminoglycoside phosphotransferase (*neoR*). This gene gives resistance to the antibiotic neomycin in bacteria and confers resistance to the aminoglycoside G418 when transfected into mammalian cells (Southern and Berg 1982). The *neoR* gene acts as a dominant marker, and can therefore be selected for in any type of cell. Koi *et al.* (1989) transfected human fibroblasts with a plasmid containing *neoR* (pSV2*neo*). The plasmid should integrate at random into any of the human chromosomes in the cell. To demonstrate this, the marked chromosomes first had to be transferred to recipient mouse A9 cells by total cell fusion (the human fibroblasts were refractory to micronuclei induction). MMCT was then used to transfer single chromosomes from the G418-resistant primary human-rodent hybrid cells to an unmarked A9 cell line. In this way monochromosome hybrids containing different tagged intact human chromosomes were constructed (chromosomes 1, 2, 5, 6, 7, 8, 10, 11, 12, 15, 18, 19, and 20). These secondary G418-resistant hybrids have proved a valuable resource in complementation experiments especially in the study of suppression of tumorigenicity



(see below). In an alternative strategy, Lugo *et al.* (1987) used a retrovirus to introduce the *neoR* gene into primary human diploid fibroblasts. Retroviral transfection has advantages over calcium phosphate-mediated transfection of plasmids in that large tandem arrays of vector DNA at the site of integration are not usually generated, and the transfection process is very efficient, making it suitable for the introduction of selectable markers into slow growing primary fibroblast cells. A complete panel of tagged mono-chromosome hybrid cell lines is now available through the repository at the Coriell Institute for Medical Research, New Jersey, USA. \*

### **1.1.3 Somatic cell hybrids as a cloning resource**

As well as enabling the localisation of hundreds of genes and anonymous DNA fragments, somatic cell hybrids can provide a cloning resource enriched for human sequences from a specific chromosomal location. Gusella *et al.* (1980) used DNA from a chromosome 11-only hybrid to create a genomic library and then used a human-specific repetitive DNA sequence as a probe to isolate those clones containing only human DNA. These clones were then localised to a specific region on chromosome 11 using a previously constructed 'panel' of DNA from deletion hybrids of chromosome 11. This strategy has since been widely adopted.

### **1.1.4 Phenotypic complementation and gene cloning**

Somatic cell hybrids provide a system in which the principles of genetic complementation used to identify genes in yeast can be applied to mammalian cells. *In vitro* expression of genes present in a particular fraction of the total genome of a mammalian cell can be studied. The phenotype observed *in vitro* of hybrids created with varying recipient cells can be correlated with the specific fragment of human DNA introduced. This allows genes to be localised according to their function.

The ability of transfected human DNA to complement the mutant phenotype of the recipient cell has already been exploited in the creation of somatic cell hybrids: those chromosome fragments which contained TK, APRT or HPRT were selected and therefore stably maintained. In this way, the sub-chromosomal location of those genes could be identified.

Perucho *et al.* (1980) extended the power of phenotypic complementation to specifically isolate and then clone the chicken TK gene, using DMGT. They transfected a mouse Ltk<sup>-</sup> cell line with a plasmid library made from genomic chicken DNA in pBR322; those hybrids in which the tk<sup>-</sup> phenotype had been complemented were selected on HAT medium. The transforming chicken TK gene was isolated from

the selected hybrids by virtue of the adjoining plasmid sequences: after digestion of the hybrid DNA with a restriction endonuclease that did not cleave within TK or pBR322 DNA, they circularised and ligated the products before transforming bacteria with them. This technique, called plasmid rescue, enabled the chicken TK gene to be recovered in bacteria where it could be readily characterised. Adapting this strategy, Seed and Arrufo (1987) constructed a human cDNA library in a transient expression vector. Upon transfection into COS cells, gene expression is driven from the vector without it integrating into the host genome. This means that the cDNA can be rescued from selected cells as an episome, without having to use the plasmid rescue technique. Using specific selection, Seed and Arrufo have isolated cDNAs encoding several cell surface antigens (Arrufo and Seed 1987 a and b). Hybrid cells expressing the desired genes were identified with specific antibodies: the episomes collected from them were used to transfect more COS cells in three rounds of enrichment until a single cDNA type was isolated, which contained the desired sequence.

#### *Cloning of DNA repair genes*

CHO cell lines mutant in genes required for DNA repair pathways have been used as recipients in somatic cell transfections to identify human genes which can correct the defects. They provide a phenotypic assay for the presence of the transfected gene by exhibiting restoration of the UV-light-sensitive DNA-repair pathway. Westerveld *et al.* (1984) describe how the human DNA repair gene ERCC1 was cloned. (ERCC stands for Excision Repair Cross Complementing rodent repair deficiency). A library of human genomic DNA was constructed in a plasmid containing *Ecogpt* which was then transfected into a CHO mutant cell line, sensitive to UV light and Mitomycin C. Transformed cells could grow on HAT medium supplemented with mycophenolic acid. Those cells which were also able to survive treatment with Mitomycin C or UV light were selected, and a genomic library was constructed from their DNA. Clones containing *Ecogpt* were identified by hybridization: this allowed enrichment for sequences which originated from the primary human cell line, and from them the gene for ERCC1 was identified. Using the same strategy, human repair genes complementing the UV sensitivity of other CHO mutants have been isolated; for example, ERCC2 by Weber *et al.* (1988) and ERCC3 by Weeda *et al.* (1990).

Weeda *et al.* (1990) also showed that ERCC3 corrects the defect in the excision repair pathway displayed in cells from patients with xeroderma pigmentosum (XP) of complementation group B (XP-B) by direct microinjection of ERCC3 cDNA into the nuclei of immortalised fibroblasts from an individual with XP. Xeroderma pigmentosum, ataxia telangiectasia (AT) and Fanconi's anaemia are representatives of

a group of human genetic disorders characterised by hypersensitivity to DNA-damaging agents. It is thought that all are due to defects in DNA repair pathways. Cell fusion experiments using cells from different patients with XP or with AT have shown cross complementation in some cases. This demonstrated genetic heterogeneity within the disorders, and subsequently at least seven complementation groups have been identified within XP (A to G) and at least four within AT (A to E): this suggests that a mutation in one of several different genes in an excision-repair pathway leads to the same overall phenotype in AT and in XP.

As we have seen, sensitivity to DNA breakage is a phenotype amenable to assay *in vitro*. Fibroblast cells from patients with these disorders can be immortalised by infection with SV40 virus or by fusion with HeLa cells, to provide an assay system for phenotypic complementation experiments comparable to those performed in CHO mutant cell lines. This has enabled several of the genes involved in these disorders to be isolated, for example the gene correcting the defect for XP A has been cloned by Tanaka *et al.* (1989) by performing two rounds of DMGT into an XP-A cell line. Interaction between genes responsible for the different complementation groups may also be studied within *in vitro* systems.

A whole chromosome 11 was identified as complementing the mutation for DNA repair in cells derived from an AT-D patient, following MMCT of the pSV2neo-tagged chromosome (Komatsu *et al.* 1990). Lambert *et al.* (1991) were able to sub-localise the region involved when a portion of chromosome 11 (11q22-23), which was translocated onto chromosome 18 during the MMCT process, could complement the defect whereas an intact chromosome 18 could not.

Transient expression from cDNA libraries in expression vectors transfected into cells from patients has also been shown to complement repair defects, and has led to cloning of genes for Fanconi's anaemia (Strathdee *et al.* 1992) and XP-C (Legerski and Peterson 1992). Cloning of DNA repair genes has been reviewed by Kaina *et al.* (1991) and their cloning and function discussed by Weeda *et al.* (1993).

### **1.1.5 Localisation of tumour-suppressor genes by phenotypic complementation**

#### *The nature of tumorigenicity*

Cell lines from tumours have been established *in vitro*. These often retain their tumorigenic properties such as the ability to grow in soft agar and the potential to cause tumours upon injection into nude mice. In 1969 Harris *et al.* showed that the high malignancy of cells from mouse ascites tumours was suppressed after they were fused with mouse A9 cells of low tumorigenicity. This observation was contrary to previous ascertainment of the 'dominant' nature of malignancy (Barski *et al.* 1961),

who observed that a fusion of malignant and non-malignant cells produced a malignant hybrid. Re-evaluation of these studies has established that tumorigenic hybrid cell lines are invariably associated with loss of chromosomes from the non-tumorigenic parent. Thus, both sets of observations are in agreement with the general concept of 'tumour suppressor' genes, described below. Later work by Cooper *et al.* (1980) and by Shih *et al.* (1981) showed that a small sub-set of tumours contain mutationally-activated oncogenes and these behave more in a truly dominant fashion, but even here the precise nature of the dominant phenotype is unclear.

Stanbridge (1976) showed that intra-specific human hybrids between malignant HeLa cells and normal diploid fibroblasts were not tumorigenic. This agreed with Harris' early results. The 'recessive' nature of malignancy in some tumours was later correlated with the loss of DNA, as outlined by Klinger (1982) and was distinct from the activation of a cellular proto-oncogene by mutation, described by Bishop (1987). (The distinction has since become less sharp with the discovery that point mutations in a single allele of some tumour suppressor genes, for example p53, can transform cells by gaining 'dominant' transforming potential (Finlay *et al.* 1989)). Identification of the DNA lost in these tumours was refined by the advent of restriction fragment length polymorphism (RFLP) analysis. This was applied by Kaebbling and Klinger (1986) when they correlated the reversion to tumorigenicity shown in intra-specific hybrids between HeLa cells and normal human fibroblasts with the loss of chromosome 11.

### *The 'two hit' hypothesis*

Retinoblastoma is a cancer of childhood. In a third of all cases of retinoblastoma, predisposition to the tumour is inherited, whereas the other two thirds of cases arise spontaneously. From the study of the incidence of retinoblastoma in both familial and sporadic forms, Knudson (1971) proposed that both copies of the wild type allele have to be lost before development of the cancer in either form. In familial cases, the first 'hit', or allele loss, is in the germ line, giving an inherited predisposition towards developing retinoblastoma when a second 'hit' occurs. The loss of the remaining wild type allele could be due to a point mutation, deletion, mitotic cross over, or a non-disjunction event. In sporadic incidence of the disease, both 'hits' have to occur within the same somatic cell of the individual.

This 'two hit' hypothesis became the model for development of cancer through the loss of function of specific genes. Using RFLP analysis, DNA loss observed in inherited cancer syndromes correlating with progression to tumorigenicity, has been characterised in order to identify specific loci, reviewed by Ponder (1988). A loss of

the second allele in tumour cells can be observed at the DNA level as a reduction to homozygosity, or loss of heterozygosity, at that locus compared to normal tissues of the individual. Although loss of heterozygosity (LOH) studies have been useful in identifying many loci involved in a progression to tumorigenicity, conclusive proof depends on developing appropriate tumour suppressor assays *in vitro* and/or *in vivo*. Transfer of DNA from wild type cells into specific tumour cell lines can be done using the techniques of somatic cell hybridization outlined above. Certain tumour cell lines can act as an assay system for chromosome fragments containing genes with tumour suppressor capabilities by demonstrating repression of their anchorage independence and ability to grow in soft agar, and/or suppression of their ability to form tumours upon injection into immunologically nude mice (reviewed by Stanbridge 1992).

Tumours are thought to arise by a multi-step process in which an accumulation of genetic changes gives rise to a series of stages in the progression from a normal cell to a metastatic, non-differentiated tumorigenic lesion. This pathway has been well characterised in colorectal carcinoma (Fearon and Vogelstein 1990). It is thought that the mutation of several tumour suppressor genes may be involved in the evolution of this cancer. By using cell lines established from different stages of other tumours in the progression to the full metastatic phenotype, it is hoped that different tumour suppressor genes may be identified by virtue of their ability to functionally complement the tumour phenotype of the different cell lines.

#### *Chromosome 11 and studies of suppression of tumorigenicity*

A single copy of a human fibroblast translocation chromosome t(11:X) has been shown to suppress the tumorigenicity of a HeLa cell line when introduced by MMCT (Saxon *et al.* 1986). The transferred chromosome contained most of chromosome 11 plus a small fragment of Xq containing the gene for HPRT, which enabled selection for those cells which had taken up the chromosome on HAT medium. An entire wild type X chromosome did not affect the tumorigenic potential of the cell line. In a similar experiment, Weissman *et al.* (1987) showed that this portion of chromosome 11 could also control the tumorigenicity of the cell line G401 derived from a Wilm's tumor (this cell line does not necessarily have the genetic rearrangements found in Wilm's tumour however). These experiments suggested that one or more genes acting as tumour suppressors are located on chromosome 11. LOH studies on the DNA from patients with Wilm's tumour, a paediatric nephroblastoma, led to the isolation of the candidate tumour suppressor gene WT1 located at 11p13 (Call *et al.* 1990 and Gessler *et al.* 1990). However, LOH at a second locus has also been observed in cases of this tumour (Reeve *et al.* 1989). Using the der(11)



monochromosome hybrid, described above, Dowdy *et al.* (1991) were able to demonstrate that DNA from 11p15.5-p14.1 was able to suppress the tumorigenic phenotype in the G401 cell line in the absence of DNA from the 11p13 region, providing evidence for a gene other than WT1 being involved in the tumorigenicity of this cell line.

The loss of tumour suppressor genes on chromosome 11 has been implicated in the development of other tumours. Oshimura *et al.* (1990) used the *neoR*-tagged 11 monochromosome hybrids (Koi *et al.* 1989, described above,) to test the ability of human chromosome 11 to suppress the tumorigenicity of various human carcinoma cell lines. They observed that hybrids made by MMCT of the chromosome 11 into a uterine cervical carcinoma and a rhabdomyosarcoma cell line were non-tumorigenic in nude mice but the same chromosome 11 transfected into a renal cell carcinoma produced a hybrid which was still tumorigenic. The introduction of a pSV2*neo*-tagged normal chromosome 11 has also been shown to complement the tumour phenotype of a cell line derived from a breast carcinoma (Negri *et al.* 1992).

#### *Suppression of tumorigenicity observed with other chromosomes*

LOH in other chromosomes has been observed in other tumours, and has led to phenotypic complementation experiments with MMCT of other chromosomes to try to identify other tumour suppressor loci.

Chromosome 6 can reduce the ability of a melanoma cell line to form clones in soft agar and to form tumours in nude mice (Trent *et al.* 1990). Chromosome 3p modulates the tumorigenicity of a renal cell carcinoma (Shimizu *et al.* 1990).

Tanaka *et al.* (1991) showed that a normal chromosome 5 or 18 could change the morphology of colon carcinoma cells, reduce their growth rate in soft agar and suppress their tumorigenicity in nude mice. This added to the evidence that loci previously identified on these chromosomes were involved in colon carcinogenesis. This work has been repeated and extended by Goyette *et al.* (1992) who also showed that an intact chromosome 17 yielded no clones at all when transfected into a colon carcinoma cell line. This effect had previously been observed by Baker *et al.* (1990) who showed that the introduction of a wild type p53 tumour suppressor gene (located on chromosome 17p) by DMGT into the same cell line produced significantly fewer clones than the introduction of a mutant p53 gene. It has recently been shown that p53 is involved in controlling cell cycle arrest (by activating the expression of the WAF1/CIP1 gene) and that it can trigger a cell's progression to apoptosis or programmed cell death (El-Deity *et al.* 1994). The region 1p36 has also been shown to suppress the tumour phenotype of colon carcinoma cells (Tanaka *et al.* 1993).

### *Strategies to refine tumour suppressor gene localisation*

Characterisation of a tumour suppressor locus on 3p has been carried out by Killary *et al.* (1992). They created an inter-specific hybrid by introducing a *neoR*-tagged human chromosome 3 into a highly malignant mouse cell line. They observed its ability to suppress tumour formation when the hybrid cells were introduced into nude mice. After prolonged culture of the hybrid cell line in the absence of G418 selection, rearrangements and loss of parts of the human chromosome occurred, and some clones regained a sensitivity to G418. Subclones derived from some of these were injected into nude mice. One line was observed to have retained its ability to suppress tumour formation. The human DNA content of this hybrid subclone was analysed by PCR and found to be limited to the region 3p21-22 thus defining a suppressor gene locus. This technique was particularly powerful in the identification of tumour suppressor loci by phenotypic complementation because inter-specific hybrids were used in the assay which enabled rapid identification of the human DNA present and also because hybrids containing a reduced human DNA content were readily generated after withdrawal of selective pressure. This final point may be due to the inherent nature of the malignant A9 clones used to make the hybrid. It is not such a straightforward strategy when applied to the study of suppression of tumorigenicity of specific human tumours where the hybrids generated are intra-specific and the transfected DNA is not readily identifiable.

Koi *et al.* (1993) have also developed a method to test the suppression activities of characterised sub fragments of human chromosomes. They randomly tagged a t(Xqter→Xq21: :11q13→11pter) human chromosome in an A9-human monochromosome hybrid with pSV2*neo*, made microcells from 18,000 pooled G418-resistant clones and fused them with A9 cells. Selection on G418 and HAT medium ensured enrichment for clones containing a *neoR*-tagged human t(11:X), with the *neoR* present at different loci in different clones. Ninety of these clones were pooled. Micronuclei made from them were  $\gamma$ -irradiated to fragment chromosomes within them before fusion with A9 cells. The human DNA in these secondary G418<sup>R</sup> A9 microcell hybrids was characterised. These clones could then be used as a resource for introducing specific human DNA fragments into different human tumours to test the suppressor capabilities of characterised loci. This is a potentially powerful, but rather labour-intensive procedure for locating human tumour suppressor genes. Controls also have to be made to eliminate the possible effect of co-transferred mouse DNA from the A9 hybrids on the phenotype of the tumours tested.

These experiments using *in vitro* phenotypic complementation assays have gone some way to elucidating the location and function of some human tumour suppressor genes, demonstrating the power of somatic cell assay systems.

## 1.2 Experimental Rationale

In any somatic cell system, it is important to be able to identify and maintain the required human DNA fragment within a hybrid. Whether it be to stably maintain a sub chromosomal region in a hybrid to be used as part of a panel of such hybrids for mapping DNA sequences from around a chosen location, or whether it be to enrich for a region for assay in a phenotypic complementation system, selectable markers are important.

We wanted to generate a series of monochromosomal hybrids in which the single human chromosome has been tagged with a selectable marker. The dominant selectable marker should be at different locations on the human chromosome in each hybrid. These hybrids could then be used as a resource for transfer of whole chromosomes or selected sub-chromosome fragments (generated by  $\gamma$  irradiation) into a cellular system to assay for genes present on those particular fragments. Additionally, by performing IFGT or CMGT and selecting for the dominant marker reduced hybrids enriched for specific chromosomal fragments can be generated. Because the progenitor hybrid contains a single human chromosome, the reduced hybrids will be useful in mapping and ordering of markers on that particular chromosome, and as a cloning resource, as described above.

We decided to introduce a *neoR* gene into the human 11-containing monochromosome hybrid J1c4 (described previously) since multiple loci on this chromosome have been strongly implicated as having an involvement in tumour development and this would allow a functional assay for tumour suppressor genes to be set up. Hybrids generated from this chromosome would also facilitate mapping studies in our group. It was proposed that human-specific repeat sequences incorporated into a targeting vector containing the dominant *neoR* tag would promote recombination with sequences in the human chromosome above those in the Chinese hamster chromosomes which make up the background of this cell line. It was also proposed that such repeat sequences may direct integration of the selectable marker into many different sites throughout the chromosome according to the distribution of repeat sequences.

We decided to test the ability of specific SINE and LINE sequences to mediate human specific integration. The origin, structure and distribution of SINEs and LINEs in human DNA is discussed below, and these sequences are compared to



similar repeat units present in Chinese hamster DNA in order to evaluate the degree of divergence and hence the likelihood of human-specific recombination events occurring. Strategies for recombination using vectors containing these repeats are discussed in section 1.6

### 1.3 Repetitive DNA Sequences in Mammalian DNA

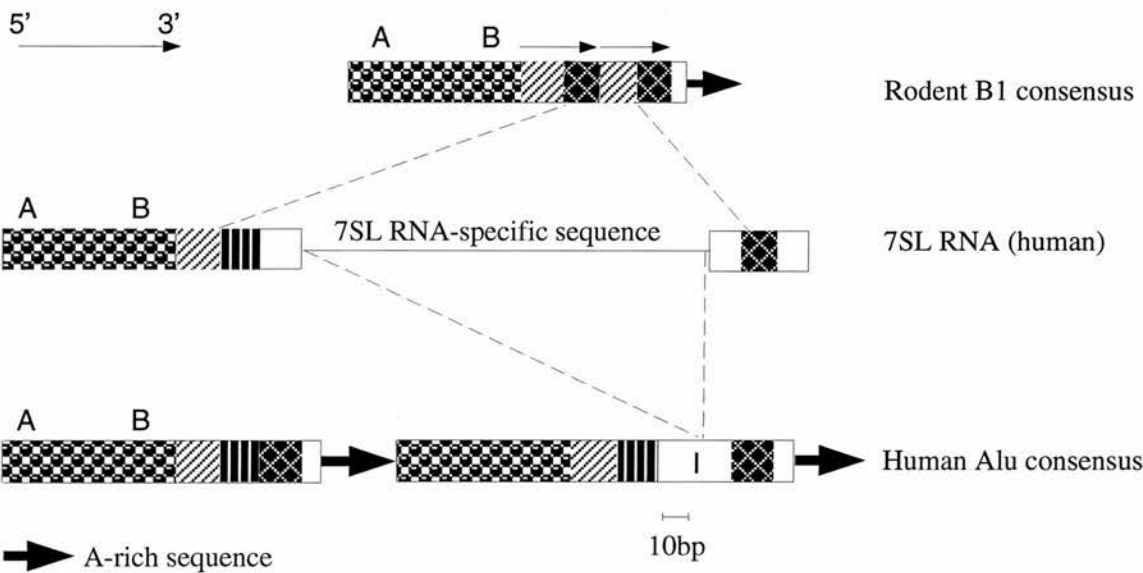
The mammalian genome contains two major families of interspersed repetitive DNA sequences: Short Interspersed Nuclear Elements (SINEs) and Long Interspersed Sequences (LINEs), first described in this way by Singer in 1982.

#### 1.3.1 The origin of SINEs: Alu and B1 families

SINEs comprise short sequences of less than 500 base pairs (bp) dispersed throughout the genomes of mammals and other higher eukaryotes. Evidence suggests that they are a class of processed pseudogene originating from the reverse transcription of small RNA molecules to form cDNA sequences which have then inserted back into the genome, a process known as retroposition (Daniels and Deininger 1985). This summary will focus on two SINE families that are thought to have originated from retroposition of the 7SL RNA gene (described below): the Alu and the B1 families.

The Alu family of SINEs is specific to primates. The consensus sequence of the family shares approximately 90% identity with the 7SL RNA gene, and contains the conserved A + B boxes which make up the internal promoter required to initiate transcription by RNA polymerase III (see figure 1.1). The 7SL RNA gene is present much earlier in evolution than the Alu repeat sequence, being conserved throughout metazoan organisms; Alu elements are restricted to primates and are also undergoing much more rapid sequence divergence than the 7SL RNA gene. This suggests that the Alu family of SINEs originated from retroposition of a processed 7SL RNA molecule (Ullu and Tschudi 1984). 7SL RNA forms part of the signal recognition particle (SP) that aides the transport of secreted proteins across the endoplasmic reticulum (ER), (Walter and Blobel 1982). It has been proposed that limited enzymatic digestion of 7SL RNA occurred; the regions which are absent from Alu sequences correspond to regions exposed on the molecule after it has adopted its secondary cruciform structure (Leigh-Brown 1984). Annealing and reverse transcription of the processed molecule gave rise to a cDNA which could then integrate into the genome as an Alu element (Ullu and Tschudi 1984).

Figure 1.1 The structural relationship of Alu, B1 and 7SL RNA sequences



The structural relationship of human 7SL RNA to the consensus sequences of human Alu and rodent B1 DNA, adapted from Ullu and Tschudi (1984). Homologous sequences are indicated by identical shading. Human Alu DNA is a head to tail dimer of two similar sequences ~130bp long. The right monomer contains an insert (I) which is not present in the left half. The B1 consensus sequence contains an internal tandem duplication of 30bp between positions 62 and 121 as indicated by the arrows. The scale is approximate.

The B1 family of SINE repeats is specific to the rodent genome, first described by Krayev *et al.* (1980). This family also has sequence similarity to the 7SL RNA gene; it is thought to have arisen by a separate processing and insertion event (see figure 1.2). The B1 repeat consensus shares some sequence identity with the Alu repeat consensus around the A + B boxes (Quentin 1989, and see figure 1.1).

There is a second major SINE family found in rodent DNA called B2, which is thought to have arisen from retroposition of rodent serine tRNA, reviewed by Deininger (1989).

### **1.3.2 The structure of SINEs : Alu and B1 elements.**

The Alu repeat is a dimer of two generic SINE units orientated in a head to tail fashion. The B1 element is a monomer. These are illustrated in figure 1.1.

### **1.3.3 Dispersal of SINEs**

SINEs are dispersed throughout mammalian genomes. The Alu family is thought to be present in at least 500,000 copies, making up approximately 5% of the human genome. The B1 family has a lower copy number of about 80,000, making up about 0.3% of the rodent genome. The B2 SINE superfamily has a similar copy number to the B1 family.

SINEs are flanked by directly repeated sequences which are not part of the family repeat unit. These are thought to arise when the repeat integrates into a new site by retroposition.

#### *Retroposition*

A proposed mechanism for retroposition as reviewed by Deininger (1992) is outlined here: RNA polymerase III transcribes the SINE sequence starting at the beginning of the element, directed by the internal promoters. It terminates at a polyA stretch in the flanking DNA. The polyU tract at the 3' end of the mRNA can pair with a short internal polyA tract to form a hairpin, which is thought to act as a primer for reverse transcription. The single stranded flanking RNA forming the loop is removed during cDNA synthesis.

#### *Integration sites of Alu repeats*

Daniels and Deininger (1985) found that sequences flanking Alu repeat elements were significantly A-rich, being ~72% d(A+T) compared to ~60% d(A+T) of the genome on average, suggesting that these elements preferentially insert into such regions of the genome. It is thought that SINE elements insert via nicks

**Figure 1.2** CLUSTAL V multiple sequence alignment of 7SL RNA (cDNA), Alu consensus left and right, and B1 consensus

```

7SL RNA cDNA      AGGC---GCCGGCGCGGTGGCGCGTGCCCTGTAGTCCCAGCTACTCGGAGGCTGAGGCTGGAGGATCGC-----
ALU LEFT          GGATCCGGCCGGCGCGGTGGCTCACGCCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCGGATCAC-----
ALU RIGHT         -----GGCGGTAGTGGCGGCGCGCTGTAGTCCCAGCTACTTTGGGAGGCTGAGGCAGGAGAAATGGC-----
B1 CONSENSUS      -----GCCGGGCATGGTGGCGCACGCCCTTTAAATCCCAGC-AC TTGGGAGGCAGAGGCAGGCGGAT-----
                  ****          ****          * ****          * ****          * ****          *

```

Sequence alignment showing the regions of identity between Alu, B1 and 7SL RNA. 7SL RNA (cDNA) bases 1-67 from Ullu *et al.* (1982), Alu consensus bases 1-71 left and 145-204 right from M. Batzer (clone pPD39 Batzer and Deininger 1991), B1 consensus from Quentin (1989).

\* indicates bases conserved across all 4 sequences.

generated during DNA replication, repair, or by random DNA damage; Lin *et al.* (1988) showed that exposure of cells to UV- light stimulated retroposition of Alu elements. Perhaps A-rich sequences are hotspots for DNA cleavage which lead to Alu insertion events. DNA Topoisomerase I introduces single-stranded nicks into DNA to relax the helix; the short direct repeats flanking Alu family members are enriched in Topo I recognition sites (Perez-Stable *et al.* 1988). There could be a specific sequence interaction between the SINE and its integration site, with the incoming DNA as either single or double stranded cDNA. Daniels and Deininger (1985) propose that the 5' end of a single stranded Alu cDNA hybridizes with a poly-T stretch exposed by nicks in the genomic DNA; this interaction provides stability, allowing cellular repair processes to join the molecules.

A study by Moyzis *et al.* (1989), using DNA reassociation, cytogenetic, computational and recombinant DNA approaches showed that the overall distribution of interspersed repeats in the human genome, is as expected for random insertion, as they show an average spacing of 3kb. However they did observe regions which were Alu-rich and those which were Alu-poor and showed that these regions corresponded to cytogenetically definable domains. Korenberg and Rykowski (1988) showed that within the human genome there is a higher density of Alu sequences in the Giemsa negative (G-) bands (or R bands). Detailed analysis of large contiguous fragments of human DNA cloned into yeast artificial chromosomes (YACs) has also demonstrated a non-random distribution of Alu repeats (Chen *et al.* 1991, Arveiler and Porteous 1992) and of L1 repeats (described below). 80% of genes reported in Human Gene Mapping 11 (1991) map to G- bands (Craig and Bickmore 1993). G- bands have also been observed to replicate early in S phase in general, compared to G+ bands which replicate late. These data suggest that Alu sequences preferentially integrate into open chromatin which is associated with replication origins and sites of transcription. Because the SINE itself contains an A-rich 3' end there is a tendency for a second element to integrate adjacent to a first; this is a possible mechanism for the formation of the dimeric Alu repeat unit (Daniels and Deininger 1985).

For newly inserted SINEs to be passed on to progeny, retroposition must occur within the germ line of the host organism. Recent Alu insertion events have been identified by family studies in humans. Wallace *et al.* (1991) identified an individual in which the repeat element had inserted into an intron of the neurofibromatosis type I gene (NF1), resulting in the deletion of a downstream exon and a shift in the reading frame of the protein: this led to the development of the disease. Neither parent of the individual showed the Alu insertion which indicated



that the retroposition event had occurred either in the germ line of one of the parents, or very early on during embryogenesis of the individual.

### **1.3.4 Isolation and characterisation of Alu repeat sequences**

The group of Houk *et al.* (1979) was one of the first to describe an ubiquitous family of repeated DNA sequences in the human genome. Previous observations had been made of 300 nucleotide duplex regions in both *Xenopus* and human DNA under the electron microscope (Deininger and Schmid 1976). Houk *et al.* wanted to characterise these regions more fully, and devised a method to isolate them from the rest of the human DNA. They denatured sheared human placental DNA and then renatured it to an association coefficient of  $C_0t = 68$ . This value had been previously determined as the  $C_0t$  at which most of the repetitive sequences in human genomic DNA have reassociated, but at which most single copy DNA remains denatured (Schmid and Deininger 1975). The reassociation coefficient is determined by the frequency at which that sequence occurs in the genome; the greater the number of copies of a sequence the faster a denatured copy will be able to reassociate with an identical sequence. The repeat sequences reassociate 50,000 times faster than single copy DNA which leads to an estimated copy number for them of >300,000 per haploid genome. The renatured DNA was treated with S1 nuclease to remove single stranded, non-annealed DNA. The remaining S1-resistant fraction was separated on a gel: about 12% of it was composed of sequences of 300bp in length, corresponding to approximately 5% of the total genome. The 300bp fraction was isolated and subjected to restriction endonuclease digestion. There was no discernable cutting with any enzyme available at that time apart from AluI which cleaved approximately 60% of the 300bp fraction into two pieces of about 170bp and 120bp. This property gave rise to the name for the major family of SINEs in human DNA : the Alu repeat. (The Alu consensus sequence has sites for many restriction enzymes since discovered including Sau3A, SmaI and BamHI). In experiments with radioactively labelled genomic DNA, they observed that the renaturation of long radioactive tracer sequences was driven by the 300bp repeat renaturation. This led them to propose that the repeats are interspersed with single copy sequences and therefore do not exist in long arrays of tandemly repeated units.

Jelinek *et al.* (1980) made sequence comparisons between the 300bp interspersed repeat found in human DNA (Alu), a similar interspersed repeat found in Chinese hamster DNA, and RNA polymerase III transcripts of cloned human DNA. They found a short region of 30bp in which nucleotides at 20 positions appeared to be conserved in all three sequences. A comparison of the sequences of 15 cloned repeat



elements from human DNA was made by Deininger *et al.* (1981). They isolated the 300bp duplex fraction of human genomic DNA in a similar strategy to that of Houk *et al.* (1979). Short linker sequences containing a BamHI restriction endonuclease recognition site were then ligated onto the blunt ends of the molecules, allowing the sequences to be cloned into the BamHI site of pBR322. Of 15 clones characterised, 10 were from the Alu family of sequences. These were named BamHI-Linked Ubiquitous Repeat (Blur) clones. Sequence comparison between the clones showed an overall similarity and revealed a dimeric formation within each clone in which there was head-to-tail arrangement of two units with considerable homology. A sequence consensus was proposed, but the data were limited.

In 1986, Bains drew up a consensus Alu repeat sequence based on 59 individual published human repeat sequences, which included the Blur clones analyzed by Deininger *et al.* (1981). He noted the A-richness of DNA flanking the repeat sequences. From his analysis, he calculated a divergence of >7% of each clone from the consensus and >12% divergence of each clone from each other.

### *SINE subfamilies*

It has subsequently been shown that Alu repeats can be divided into subfamilies, each having a consensus from which sequence divergence is relatively small.

Initially, a division into only two major subfamilies was proposed for Alu elements (Jurka and Smith 1988). Since then, several subfamilies of Alu repeats have been identified. They are thought to have originally arisen from very few retropositionally competent master genes, with evolution occurring through a succession of subfamilies rather than continual exponential expansion. Different subfamilies have arisen at different times in evolution, with the more divergent families having arisen earlier (Britten *et al.*, 1988, Matera *et al.* 1990a). A relatively young subfamily has been identified which makes up 10% of Alu repeats in human DNA (Britten *et al.* 1988), and is known as the Precise subfamily.

A consensus sequence and a similar subfamily structure has been drawn up for B1 repeats in rodent genomes by Quentin (1989).

### *Active subfamilies of Alu and B1 repeat sequences*

Some Alu repeats are dimorphic within the human population, being present or absent at specific loci in different individuals. These recently inserted sequences are the most conserved, with a few notable exceptions (discussed later), and a sub group which share 5 mutations relative to the Precise subfamily has been identified. This sub-group was named PV for Predicted Variant by Matera *et al.* (1990 a and b),

and HS (as it was believed to be Human Specific) by Batzer and Deininger (1991), identified from separate experiments. Through selective oligonucleotide hybridization both groups were able to distinguish members of the PV or HS group above the background of other Alu repeat sequences. Matera *et al.* (1990a) also observed mRNA transcripts in various tissues by Northern blot hybridization, corresponding to this subfamily, thus showing that the PV subgroup is transcriptionally as well as transpositionally active. A consensus sequence was drawn up for PV (Matera *et al.* 1990a) which was equivalent to the one for HS (Batzer and Deininger 1991). The sequences of the newly inserted Alu discovered in an NF1 gene by Wallace *et al.* in 1991, differed from the PV consensus by only 2 nucleotides, confirming that PV is an active Alu repeat subfamily.

A subfamily of B1s which can exist as small cytoplasmic RNAs called scB1 has been characterised by Maraia (1991). These scB1 sequences are very similar to the B1 consensus but have undergone RNA processing at their 3' end and lack a poly-U tail, which means that they are unlikely to be intermediates for retroposition. A possible function for these molecules is discussed below.

### 1.3.5 Evolution of Alu Subfamilies

Studies of the evolution of Alu subfamilies and particularly the spread of members of the PV subgroup have led to two different theories for the pattern of their dispersal.

#### *A Single Master Gene*

From the group of P. Deininger, Batzer *et al.* (1990) and Shen *et al.* (1991) propose that Alu sequences have originated from a single master gene and that at any one time, only one Alu sequence at a single allele in the genome is transcriptionally active and capable of retroposition (reviewed by Deininger *et al.* 1992). They explain the formation of the Alu subfamily structure by allelic variation at a single master gene locus. They propose that of the many Alu sequences which may arise from retrotransposition, 99% integrate into a locus where they are transcriptionally and/or retropositionally silenced. The high homology between recently inserted Alu sequences suggests that they have all arisen from a single locus and not from several loci in which different random point mutations would have accumulated. They propose that older subfamilies have arisen in a similar fashion throughout primate evolution; each from a single master gene which produced many copies which were then subject to random mutation. When a mutation occurred in the master gene, all subsequent copies would contain this mutation, and thus a new subfamily would be

formed, with no further transcription of the old master gene sequence. The groups led by Deininger invoke this hypothesis to explain the observations of a punctuate mutation rate through Alu evolution. They also propose that the amplification rate of Alu sequences that was in its peak early in primate evolution has diminished since then due to accumulated mutations in the master gene which have resulted in it being less efficiently amplified.

### *Multiple Master Genes*

Papers with evidence opposing this hypothesis have come from several groups including Matera *et al.* (1990a and b), Stoppa-Lyonnet *et al.* (1990), Leeflang *et al.* (1992), and Jerka (1993), reviewed by Schmid and Maraia (1992). Stoppa-Lyonnet *et al.* (1990) discovered that a member of the subfamily Precise, believed to be older than the PV subgroup was polymorphic in the human population, having inserted into the C1 inhibitor locus after human-ape divergence. Shen *et al.* explained this polymorphism as having arisen from the Precise founder gene before it underwent mutation to become the founder for the PV subgroup, presupposing that the PV subgroup is human specific. In direct contradiction to this theory is the observation by Matera *et al.* (1990a) that members of the PV family are present in DNA from gorilla, chimpanzee, and bonnet monkey. Leeflang *et al.* (1992) detected PV sequences in chimpanzee DNA and were able to propose that if there was a single founder gene for the PV subfamily, it must have existed before the human-ape divergence. There must therefore have been multiple Alu source genes active contemporaneously. A multiple master gene hypothesis explains the rapid expansion in numbers of PV elements in human DNA since their divergence, and their high degree of polymorphism. It also explains the origin of the novel Alu insertion in the cholinesterase gene reported by Muratani *et al.* (1991), which shows higher sequence homology with Precise than with the PV subgroup. Jerka (1993) proposes that this particular sequence defines a new sub group of Alu sequences and names it AluSb-2, where AluSb = Precise and AluSb-1 = PV.

The evidence is therefore in favour of a model in which the evolution of Alu repeat elements is controlled by relatively few but simultaneously active master genes.

There is also evidence for the involvement of multiple active master genes in the evolution of B1 subfamilies (Quentin 1989).

### **1.3.6 Inactivation of Alu Sequences**

All groups agree that only a very small proportion of Alu sequences are transcriptionally and/or retropositionally active. The local environment of an inserted

Alu element is thought to determine its activity. An upstream enhancer sequence or transcription from a nearby gene may be required for transcription to be initiated from the RNA Pol III promoter within the Alu element (Paulson and Schmid 1986). Downstream sequences could determine the ability of a transcript to self-prime and therefore affect its ability to form a cDNA which can re-insert into the genome. Methylation of CpG dinucleotides and/or accumulated mutations at these points could affect transcriptional activity ('young' Alu sequences contain up to 9% CpG dinucleotides, the average for the genome is 1%). Perhaps the local chromatin structure plays a large part in determining whether or not an Alu is transcribed. The availability of reverse transcriptase in the cell will also be important. Factors influencing Alu inactivation are reviewed by Schmid and Maraia (1992).

### **1.3.7 Is there a function for Alu and B1 repeat sequences?**

The most obvious answer to this question is that SINEs are simply 'selfish' DNA with no function other than to propagate within the genome of the host. Any other function may be incidental. Their presence in eukaryotic genomes is not essential, as some organisms (for example *Fugu rubripes* (Brenner *et al.* 1993) and *Arabidopsis thaliana* (Anderson 1991)) have very few, if any, SINEs in their DNA, so it could be postulated that there has been limited evolutionary selection either for or against retaining these sequences.

However, short interspersed nuclear elements are widely dispersed and present in large copy numbers and so must have some effect incidental or otherwise on the structure and function of the genomes in which they reside. They may play a role in the organization of the structure or methylation status of chromatin by virtue of their high GC content, or perhaps their short polyA stretches can provide transcriptional termination signals where they might otherwise be lacking. The transcriptional and retropositional competence of Alu and B1 RNA has potential effects on gene disruption through insertional mutagenesis (Wallace *et al.* 1991) and genomic instability. Also, there is evidence that Alu sequences might provide hotspots for illegitimate or legitimate recombination within the host genome; they are frequently found at translocation breakpoints for example (discussed in more detail below). There has been recent speculation as to the possible function of a subset of Alu transcripts observed to accumulate in the cytoplasm of mammalian cells (Matera *et al.* 1990a). Homology of these transcripts with scB1 RNA was observed and they were termed scAlu by Maraia *et al.* (1993). This group proposes that scAlu RNA arises by post-transcriptional processing of full length active Alu mRNA to produce an RNA molecule approximately 120 nucleotides long, which lacks a poly-U 3' tail and is



equivalent to a monomeric SINE unit derived from the 5' end of the Alu repeat. Chang and Maraia (1993) have shown that scB1 and scAlu have conserved a secondary structure motif despite substantial sequences divergence through evolution. It has been proposed (Chang and Maraia 1993, Maraia *et al.* 1993) that some active Alu and B1 elements produce transcripts which can undergo 3' processing to form scAlu and scB1 molecules. This processing allows them to associate with a cellular protein to form part of ribonucleoprotein (RNP) particles which accumulate in the cytoplasm and may perform a specific cellular function.

### 1.3.8 LINES

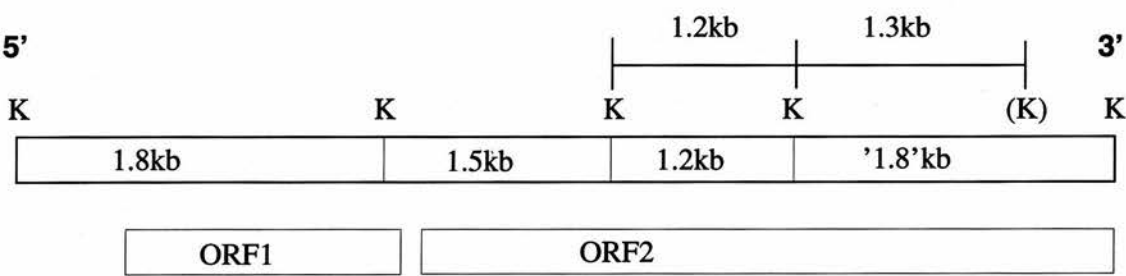
LINES are the other major super family of interspersed repetitive elements in mammalian DNA. They were first identified in primates as a series of characteristic bands seen on an ethidium bromide-stained agarose gel of genomic DNA digested with KpnI and subjected to electrophoresis (Maio *et al.* 1981). Reassociation studies revealed that they exist as interspersed elements, with an overall copy number of 50,000-100,000 per haploid genome, making up approximately 5% of the human genome (reviewed in Hutchinson *et al.* 1989).

A KpnI repeat at the 3' end of the human  $\beta$ -globin locus was characterised by Shafit-Zagardo *et al.* (1982). A repeat unit of ~6.4kb was seen as comprising four KpnI fragments (see figure 1.3). Only about 10% of these repeat units are full length; the majority of sequences are truncated at their 5' end (Grimaldi *et al.* 1984). The copy number of the repeat sequence varies from 100,000 for most 3' sequences to 10,000 for 5' sequences.

### 1.3.9 The Line 1 (L1) repeat family

A repeat unit > 5kb in length was recognised as being also conserved in rodents, originally termed the BamHI or MIF-1 repeat family (Singer *et al.* 1983). This super family of LINES showing conservation throughout mammalia has been named Line1 (L1) (Voliva *et al.* 1984). L1 repeats are divided according to the species from which they originate for example, L1-H from humans, L1M-d from mouse and L1Cg from Chinese hamster (Miles and Meuth 1989). A consensus sequence for L1-H has been drawn up (Scott *et al.* 1987). Subfamilies can be assembled by virtue of their divergence from the consensus (Scott *et al.* 1987, Rogan *et al.* 1987), suggesting that L1s have derived from a small number of active genes. A comparison of consensus sequences for specific species subfamilies reveals that through concerted evolution, there is greater sequence divergence interspecies than intraspecies (Burton *et al.* 1986).

Figure 1.3 A Line 1 repeat element



K = KpnI restriction site

A representation of a full length L1 element, showing the fragments used in the targeting vector L(1.2), L(1.3) and L('1.8'), modified from Shafit-Zagardo *et al.* (1982).



Two open reading frames (ORFs) have been identified which lie towards the 3' end of the L1 repeat (see Fig 1.3), and which are within the region conserved between species (Burton *et al.* 1986).

### 1.3.10 L1 repeats are retroposons

Several of the characteristics of LINEs suggest that they are dispersed through the genome by retroposition. L1 elements are flanked by short direct repeats similar to those found at the insertion sites of SINEs (Voliva *et al.* 1984). Active L1 elements have been identified in human individuals where their insertion has disrupted a gene leading to the sporadic incidence of a disease, for example in the Factor VIII gene (Kazaian *et al.* 1988 and Dombrowski *et al.* 1991, 1993) and the APC gene (Miki *et al.* 1992), suggesting that these elements are mobile. Transcription of L1 repeats by RNA polymerase II had previously been observed in human cells (Shafit-Zagardo *et al.* 1983) and full length (~6.5kb) poly-adenylated transcripts have been observed in the cytoplasm of human teratocarcinoma cells (Skowronski *et al.* 1988). Truncated L1 sequences may arise by transcription initiating between the ORFs, or by incomplete reverse transcription of the full length mRNA (Scott *et al.* 1987). Translation of transcripts encoded by the L1 repeat has also been observed *in vitro* and in human cells. ORF1 encodes a ~40kda protein whose function is yet unknown (Leibold *et al.* 1990); ORF2 has been shown to encode a protein of ~150kda with reverse transcriptase function (Mathias *et al.* 1991).

Taken together, their characteristics provide compelling evidence that the LINE sequences are dispersed through mammalian genomes through an RNA intermediate, and that they encode proteins, at least one of which facilitates their own propagation, that being reverse transcriptase. Transcription is thought to be mediated by RNA polymerase II, initiated from a promoter lying immediately 5' to the L1 repeat unit (reviewed in Hutchinson *et al.* 1989). The presence or absence of this promoter will determine the activity of that particular LINE. It is proposed that only relatively few L1 elements are active at any one time. \*

### 1.3.11 Distribution of L1-H elements

The distribution of L1 sequences in the human genome has been observed to be complementary to that of Alu sequences, with Alu -rich regions being L1 poor and vice versa. Alu and L1 elements separate into different fractions after density gradient centrifugation of human genomic DNA, (Soriano *et al.* 1983). Evidence for a non-uniform and complementary distribution of Alu and L1 repeats has also come from detailed analysis of several representative human DNA fragments; for example, in

yeast artificial chromosome (YAC) clones (Chen *et al.* 1991, Arveiler and Porteous 1992) and Not I restriction fragments (Sainz *et al.* 1992).

In their original studies of the cytogenetic distribution of repeat sequences, Korenberg and Rykowski (1988) showed that L1 were more prevalent in G+ bands, being excluded from the areas of strong Alu hybridization, the G- bands (reviewed in Bickmore and Sumner 1989).

## 1.4 SINE and LINE sequences used in the targeting vector

### 1.4.1 Blur8

The SINE included in the targeting vectors was the Blur8 Alu sequence (as described above). This sequence shows 84% similarity to the PV (HS) consensus (Batzner *et al.* 1991), and 87% similarity to the overall consensus proposed by Bains (1986). Divergence of Alu sequences from the overall consensus is >7%, with each Alu sequence varying by ~12% from any other Alu element (Bains 1986). Blur8 shows 45% divergence from the rodent B1 consensus given by Quentin (1989), and approximately 40% divergence from a Chinese hamster ovary 'Alu equivalent' sequence described by Haynes *et al.* (1981).

### 1.4.2 L(1.2), L(1.8) and L(1.3)

The LINE sequences included in the targeting vectors were L(1.8), L(1.3) and L(1.2). L(1.8) and L(1.2) were cloned from the 5' end of the  $\beta$ -globin locus (see Shafit-Zagardo *et al.* 1982 described above). L(1.3) was isolated from a human genomic DNA KpnI library, and cross hybridises with sequences from the 3' end of the L1 element. It was a gift from E. Slorach (Slorach and Brookes, unpublished).

The divergence of the sequence of 3' fragments from the overall consensus of the LINE is <5% as this region is contained within ORF2 of the LINE (Scott *et al.* 1987). The interspecies divergence in this region is higher, being in the region of 25-45% overall (Scott *et al.* 1987).

### 1.4.3 Repeat sequences and targeting vectors

The characteristics of the repeat elements incorporated into the targeting vectors are summarised in Table 1.1. With the chosen repeat elements, we proposed that the human intraspecies divergence was sufficiently less than the hamster/human sequence divergence to favour repeat-mediated recombination with human sequences over that with Chinese hamster sequences. Due to the complementary nature of the distribution of these two types of repeat in the human genome, we decided that it would be advantageous to incorporate both these elements into different vectors for targeting,

Table 1.1 Characteristics of Alu and Line1 human-specific repeat families

Repeat element	Family of repeat	Prevalence of family members	Conservation within family	Rodent/ human homology
BLUR8	SINE: Alu	>500,000 per genome >12,500 per chr. 11 Predominantly in R-bands	each repeat varies ~12% from consensus	55% of bases are conserved over 138bp
L(1.2) L(1.8) L(1.4) <sub>3</sub>	LINE: L1-H	50,000-100,000 per genome 1,000-2,500 per chr. 11 Predominantly in G-bands	each repeat varies < 5% from consensus	up to 65% overall homology within ORFs

to achieve an even distribution of dominant selectable markers throughout chromosome 11.

The repeat sequences were incorporated in a targeting vector designed to maximise recovery of recombination events between the repeat elements in the vector and Alu and L1 sequences in the human chromosome by a 'positive-negative' selection procedure. A discussion of recombination in mammalian cells and the design of targeting vectors is made below.

## 1.5 Recombination in Mammalian cells

### 1.5.1 Models for homologous recombination

The pairing of two identical sequences followed by DNA breakage and exchange of DNA strands is known as homologous recombination (HR). This phenomenon occurs at a high frequency in *Saccharomyces cerevisiae*, and has been widely studied in this organism (reviewed briefly by Hastings 1988). A double strand break repair (dsbr) model has been proposed as the mechanism of homologous recombination in yeast and other lower order fungi (Szostak *et al.* 1983), shown in Fig. 1.4. In this model, a free 3' end created by the action of exonucleases at a double stranded gap in a DNA molecule invades a homologous duplex. A loop of DNA is displaced (D-loop) and this is enlarged by 3'→5' DNA synthesis until it can anneal with complementary sequences on the other free 3' end. Repair synthesis from the second strand occurs and completes the process of gap repair. Two Holliday junctions are formed which are resolved by cutting of the inner or the outer strands leading to four possible final products: two crossover and two non-crossover. It is a conservative recombination mechanism, in which all the sequences present before the recombination are present after the event. This model appeared to fit the meiotic segregation products observed in the lower order fungi better than the model previously proposed by Meselson and Radding (1975). In the latter model (shown in Fig. 1.5), only a single strand nick is required to initiate recombination. The free 3' end invades a homologous duplex, and the mechanism is propelled by 3'→5' DNA synthesis from the 5' side of the nick. The displaced D-loop is degraded by the exonuclease, having no sequence to form a stabilising duplex with. Branch migration brings the 5' and 3' single stranded ends together so that they can be ligated, forming a single Holliday junction. Resolution of this junction can again bring crossover or non-crossover products.

Figure 1.4 Double stranded break repair (dsbr) model  
(taken from Szostak et al. 1983)

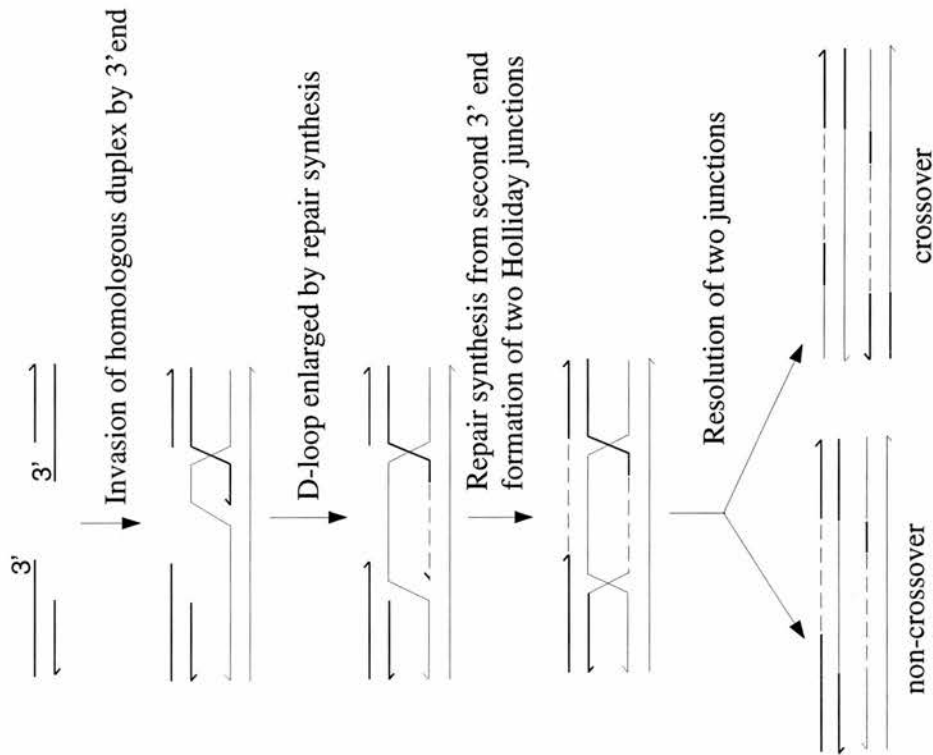
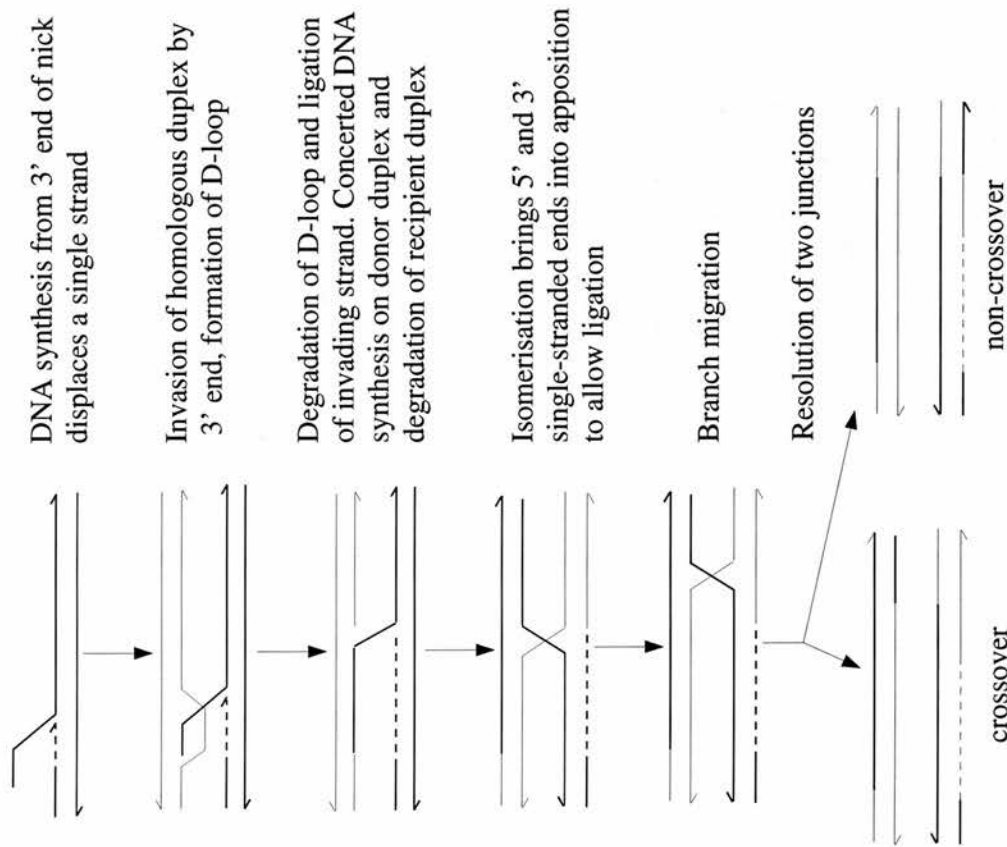


Figure 1.5 The Meselson- Radding model  
(taken from Szostak et al. 1983)



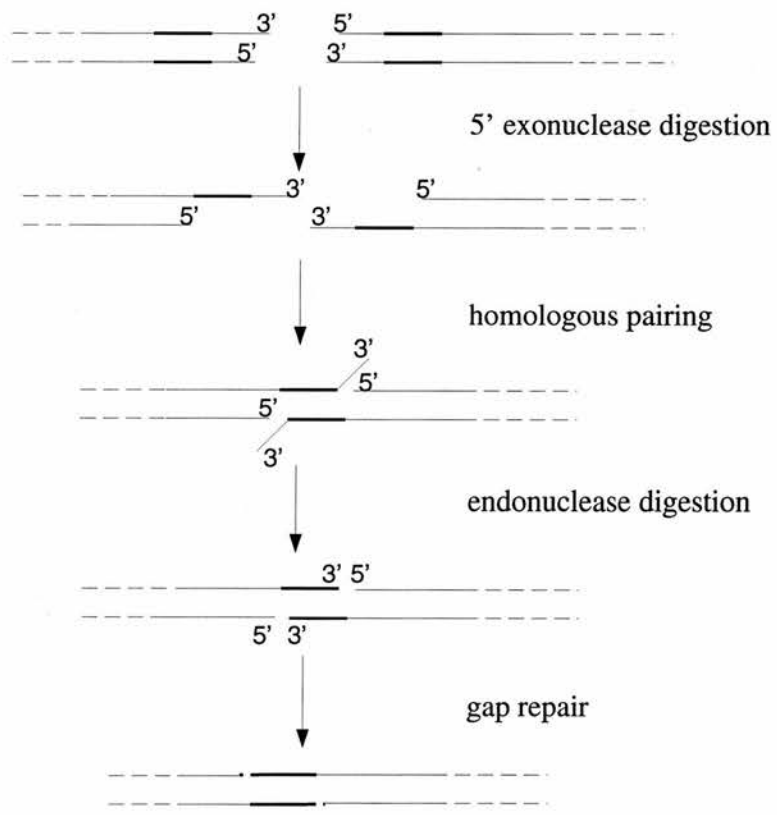
### 1.5.2 Extra chromosomal recombination in mammalian cells

In a first attempt to elucidate some of the mechanism(s) of recombination in mammalian cells a simplified system using extra chromosomal (EC) plasmid or viral DNA molecules was adopted. To test for homologous recombination (HR) using this system, two molecules co-transfected into mammalian cells are designed to be defective in a specific gene but to carry different mutations. HR is observed by the generation of wild type virus particles (resulting in lytic plaques) or by the production of a functional gene conferring resistance to selective medium. Using this system, Anderson *et al.* (1984 and 1986) noticed that recombination could indeed occur between plasmids co-transferred into mammalian cells to generate a functional selectable marker and that short regions of homology between the plasmids were present at the point of recombination. Linearisation of injected plasmids was observed to greatly stimulate homologous recombination and that the plasmids recombined through the ends of the molecules (Folger *et al.* 1982). Confirmation that HR could occur in mammalian cells came from experiments in which bacteriophage  $\lambda$  or plasmid molecules containing two different mutant tk genes introduced into Ltk<sup>-</sup> mouse cells gave a tk<sup>+</sup> phenotype which must have arisen through HR (Small & Scangos 1983). The frequency of extra-chromosomal recombination (ECR) expressed as a fraction of the number of cells giving G418 resistance when transfected with an equal amount of wild type *neoR*, was observed to be as high as 7%, which was enhanced 10 fold by linearisation of one of the plasmids (Kucherlapati *et al.* 1984). This and other data led Bollag *et al.* (1989) to postulate that HR between EC substrates in mammalian cells could occur as frequently as once per cell.

The development of these ECR systems meant that models for recombination in mammalian cells could be proposed and tested. Lin *et al.* (1984) proposed a single strand (ss) annealing model for ECR in mammalian cells (shown in Fig. 1.6). In this model, single-strand exonucleases degrade the ends of a linear DNA molecule to an extent which exposes complementary strands of the two defective genes in both molecules. The complementary regions then pair, unpaired regions of DNA are removed from the junction and a gap repair process then reconstructs an intact gene. Other groups have provided more evidence for this non-conservative recombination mechanism between EC substrates in mammalian cells. Wake *et al.* (1985) proposed that the DNA is unwound by a helicase to produce the single strands involved in the recombination instead of the exonuclease 'stripping' advocated in the original model. These ends are then removed by endonucleases and the gaps repaired as before. Anderson and Eliason (1986) extended the ss annealing model, providing results



Figure 1.6 Single stranded annealing model for recombination  
(taken from Lin et al. 1984)



which suggest that HR of linear EC molecules occurs by the non-conservative joining of free ends in a terminal pairing mechanism. Lin *et al.* (1987, 1990a and b) provided more evidence for the ss annealing model through the study of double-stranded (ds) breaks, exonuclease action and the length of homology between the substrates.

Other groups have, however, provided evidence that the dsbr model proposed for recombination in yeast fits the results obtained for extra chromosomal HR. Song *et al.* (1985) evoked the dsbr model to explain the persistent observation that a ds break within a region of homology stimulates HR. They suggested that ds breaks initiate the recombination event and promote a gene conversion event. Brenner *et al.* (1985) and Jessberger and Berg (1991) also have results which support the dsbr model for HR of extra chromosomal substrates. Brenner *et al.* presented data showing that regions of homology flanking a ds break enhance the frequency of extra chromosomal HR and that this fits a model in which homologous pairing of the substrates occurs prior to a strand invasion initiated by the ds free end. The ss annealing model does not readily explain the fact that recombination frequency is increased by flanking homology that is separated from the region in which recombination has occurred by a ds break. Following the ss annealing model, exonuclease activity will remove flanking homology from one side of the break and therefore this DNA can play no part in stimulating HR. Jessberger and Berg (1991) examined the products of ECR in a mammalian *in vitro* system using labelled plasmids as substrates for recombination. From a detailed analysis of the products they could trace the fate of the substrate molecules involved and invoked a dsbr model to account for their observations.

### **1.5.3 Recombination between an endogenous locus and transfected DNA**

The studies described above did not address the phenomenon of recombination of exogenous DNA with endogenous chromosomal sequences. Lin *et al.* (1985) were among the first to show that recombination can occur between DNA introduced into mouse L cells and homologous chromosomal sequences. They introduced a plasmid containing tk $\Delta$ 3' into mouse Ltk $\Delta$ 5' cells and observed that an intact tk gene was generated by functional assays and Southern blotting. This event was very rare, however, occurring  $1 \times 10^5$  times less often than integration of the plasmid at random. Thomas *et al.* (1986) introduced a mutant *neoR* gene into mouse cells at random and then corrected the mutation by HR using a plasmid containing *neoR* with a different mutation, to produce cells that were now resistant to G418. Using the technique of microinjection to introduce the plasmid into individual cells they obtained an efficiency of HR events of 1 in  $10^3$  cells receiving DNA.

Targeting of endogenous loci was extended to genes not directly selectable *in vitro* by Smithies *et al.* (1985) when they were able to introduce an exogenous *neoR* marker into the  $\beta$ -globin gene using a plasmid containing  $\beta$ -globin sequences. The recombination occurred exactly as was predicted by a conservative recombination event through regions of homology. When the  $\beta$ -globin gene was active, targeted cells could be selected by their resistance to G418. Again, frequencies of HR were fairly low, with only 1 in  $10^3$  cells transformed by electroporation containing the targeted event.

HR of extra-chromosomal substrates can occur with as little as 25bp of homology and with longer sequences it can tolerate relatively high levels of mis-match between the sequences (reviewed by Bollag *et al.* 1989). HR with endogenous loci seems to require greater than 200bp of homology and its frequency is significantly reduced by heterologies (te Riele *et al.* 1992, Deng and Capecchi 1992). It has also been shown that the frequency of intra chromosomal recombination in mammalian cells is dependent upon the length of uninterrupted homology between the sequences involved (Waldman and Liskay 1988). Two sequences containing mismatches but maintaining 232bp of uninterrupted homology recombined 20 fold more efficiently than did sequences in which mismatches reduced the length of uninterrupted homology to 134bp. Since a mechanism involving gene conversion is thought to be involved in recombination events involving two chromosomal loci (Bollag *et al.* 1989), i.e. intra chromosomal recombination, the same requirements may apply to targeted recombination events. HR between extra chromosomal substrates also occurs at a higher frequency than it does between an introduced plasmid and its chromosomal target.

It is likely, then, that the mechanisms involved in homologous recombination between two exogenous substrates and those involving endogenous substrates are different at least in part, although some common intermediates and proteins may be involved.

#### **1.5.4 Illegitimate recombination**

Random or illegitimate recombination (IR) occurs far more frequently than HR in mammalian chromosomes. Illegitimate recombination between chromosomes or between a plasmid and an endogenous sequence can occur at very high frequencies in mammalian cells. A general mechanism has been proposed for IR in mammalian cells whereby recombination is stimulated by free DNA ends and occurs when the ends pair, mediated by 1-5 nucleotides of homology: a terminal pairing mechanism (reviewed by Roth and Wilson (1988)). Free DNA ends can arise by several

mechanisms in normal mammalian cells, for instance during DNA replication or during VDJ recombination during maturation of the immune system. In fact IR is used by the immune system to generate an extra degree of diversity in developing lymphocytes and T-cell receptor genes (Lewis *et al.* 1985). When IR occurs inappropriately it can give rise to chromosome translocations or deletions. It has been proposed that a ss DNA break would predispose to the generation of small, base pair deletions, whereas a ds break could generate large chromosomal deletions. Due to the large amount of non-essential DNA in the mammalian genome high frequencies of IR can be tolerated, and IR has been important in generating diversity during the evolution of genes. The frequency of IR in yeast however, is low. In *S. cerevisiae* for example, recombination through regions of homology occurs 10 times more frequently than does random recombination. The relatively low rate of HR in mammalian cells is not explained by the target frequency due to the comparative size of the genomes. Thomas *et al.* 1986 found no difference in targeting frequency when they increased either the quantity of plasmid DNA injected into a mouse cell line or when the number of target sites was increased from 1 to 5 at the chromosomal locus. Experiments by Zeng and Wilson (1990) demonstrated no differences in the frequencies of HR when comparing the targeting of a normal CHO cell line carrying a single gene at the DHFR locus or a cell line with amplification of up to 800 copies of the DHFR sequence. Wallenburg *et al.* (1987) failed to detect any HR events when a rat SINE was introduced into a rat cell line. Perhaps a difference in the amounts of enzymes and other proteins which mediate HR accounts for the difference in frequency of HR between yeast and mammalian cells. Enzymes catalysing IR in mammalian cells are perhaps able to act more efficiently than those promoting HR and therefore IR is the predominating fate for free DNA ends in those cells. In an interesting case, Buerstedde and Takeda (1991) reported a predominance of targeted to non-targeted events with constructs transfected into a chicken B cell line. Four different loci were targeted, one of which was highly transcribed and another transcribed at very low levels, if at all in the cell line. All four loci showed at least 50% and one showed 100% of selected clones to have targeted events. The ratio of targeted to non-targeted events dropped by an order of magnitude when the same constructs were transfected into non B cell lines. These experiments suggested that the cell line analysed over-expressed an enzymatic activity that facilitates HR, supporting the theory that recombination machinery determines the frequency of homologous recombination in mammalian cells.

Overall then, this suggests that random and homologous recombination occur by different mechanisms in eukaryotic cells. Perhaps random recombination occurs in a similar way as HR between extra chromosomal substrates, where the requirement

for homology can be very low. The question remains as to whether chromosomal HR in mammalian cells occurs by the same mechanism as in yeast. Some evidence supporting a dsbr model of HR in mammalian cells has come from work examining the factors which influence HR and is discussed below.

### **1.5.5 Gene targeting in mammalian cells**

The ability to precisely alter genes within mammalian cells has been recognised as a powerful tool in the study of gene function. Gene targeting events engineered in pluripotent murine embryonic stem (ES) cells by HR with an exogenous plasmid has rapidly become a key methodology in mammalian genetics (Thomas and Capecchi 1987, Doestchman *et al.* 1987, reviewed by Capecchi 1989 a and b). ES cells can retain their pluripotency after genetic manipulation *in vitro* and following injection into mouse blastocysts may contribute to the germ line of an animal which develops from the blastocyst (Thompson *et al.* 1987). In this way, transgenic animals can be produced in which an engineered mutation introduced precisely by an exogenous plasmid is contained within every cell of that animal. The effect of the absence of a gene product, or the presence of a gene product with a precise mutation can therefore be monitored not only within a cell culture, but can also be observed *in vivo* as it is manifest on the phenotype of a whole animal.

Recombination mechanisms have been investigated in relation to gene targeting in order to find ways of decreasing the frequency of random integrations relative to HR events in mammalian cells. Valancius and Smithies (1991) found that the dsbr model for HR could explain their observations of ds gap repair of an incoming plasmid by a homologous endogenous sequence. Pennington and Wilson (1991) showed that unlike HR of extra chromosomal substrates, gene targeting in mammalian cells occurs by a conservative pathway. This therefore supports the dsbr model rather than a ss annealing or terminal pairing model for HR between an incoming plasmid and its chromosomal target. Plasmids designed to introduce selectable markers by homologously recombining with a target locus are divided into two main types which are outlined below.

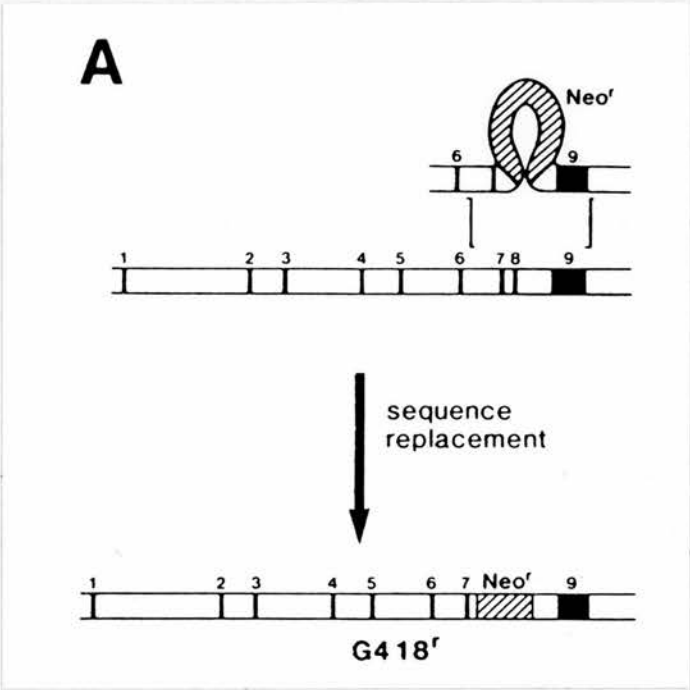
### **1.5.6 Insertion and replacement targeting vectors**

Insertion vectors contain a single region of homology, in the middle of which is the point of linearisation of the plasmid. These vectors integrate through a single cross over event which results in the insertion of the entire plasmid sequence into the targeted locus (see Fig. 1.7 b). Replacement vectors, on the other hand, are designed to replace the endogenous sequence by that on the incoming plasmid. They may

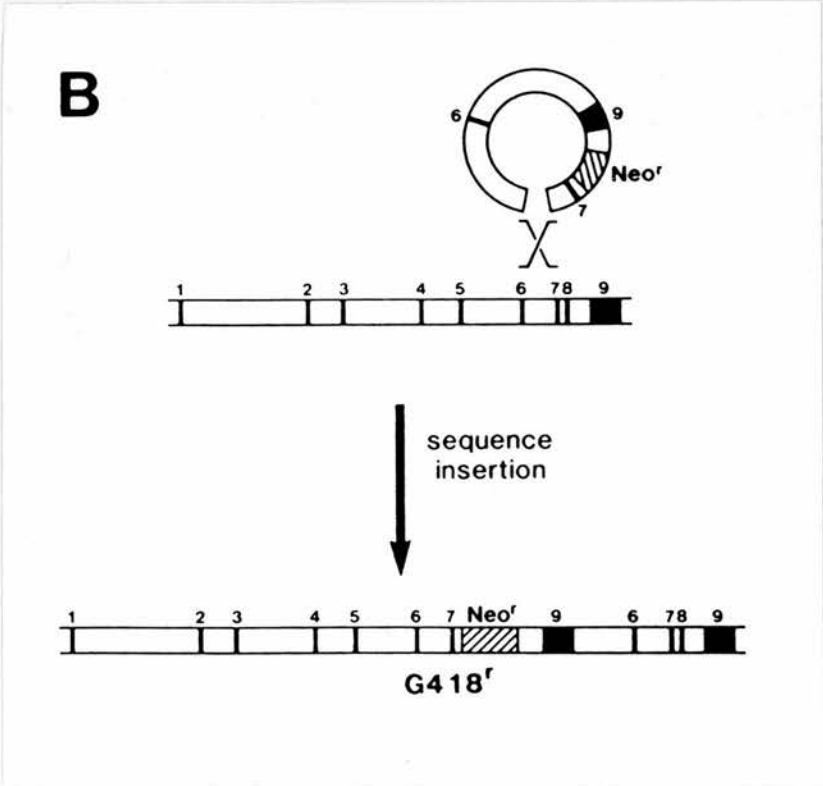


Figure 1.7 Recombination mechanisms for different types of targeting construct

Replacement  
vector



Insertion  
vector





incorporate a negative selection marker which allows enrichment for integration events which have occurred through two crossovers mediated by two separate regions of homology present in the vector (see Fig. 1.7a), described by Mansour *et al.* (1988).

There is some debate over the comparative efficiencies of the two types of vector (discussed by Hasty *et al.* 1991b, Dickinson *et al.* 1993, Thomas *et al.* 1992, Deng and Capecchi 1992, and Deng *et al.* 1993). What emerges is that rather than the overall design of the vector, the frequency of HR seems to depend more on the properties of the recombining sequences. These properties include the total length of homology present in the vector and the length of homology on each side of the ds break; heterologies between vector DNA and the target DNA i.e. whether or not the two regions of DNA are isogenic; and the intrinsic recombinogenic status of the targeted locus, affected for example, by its transcriptional activity. Hasty *et al.* (1991 a) found that a total of >1.9kb of homology was required in a replacement-type targeting vector to generate targeted clones. They also showed that the efficiency of gene targeting increases 200 fold with an increase in the length of homology from 1.3kb to 6.8 kb for both replacement and insertion vectors. However, targeting occurred with as little as 472bp of homology on one of the arms, demonstrating that only a short region of homology is required for the resolution of a Holliday junction. Deng and Capecchi (1992) showed that vectors containing isogenic DNA targeted 4-5 times more efficiently than did the corresponding vectors containing non-isogenic DNA, and te Riele *et al.* (1992) discovered a 20 fold increase in efficiency using isogenic compared to non-isogenic DNA.

Terminal heterologies may or may not be incorporated into the site of integration. Hasty *et al.* (1992) have observed some events where short heterologous vector sequences (13 bp) are present at the integration site using an insertion vector. Dorin *et al.* (1992) provide evidence for precise recombination of an insertion vector, where the very ends of the vector remain intact, to the extent that the restriction endonuclease site at the point of linearisation of the plasmid is preserved at the site of integration. Kumar and Simons (1993) found that long heterologous sequences (~2kb) were removed from the ends of insertion vectors upon HR, with some reduction in efficiency but not in fidelity of gene targeting.

### **1.5.7 Mechanisms of gene targeting**

From work published so far, it appears that the search for a homologous sequence during gene targeting does not absolutely require homology at the very end of the vector. The recombination process, however, does require homology and unless the heterologous ends are very short (<13 bp) then the unmatched sequences are removed

(probably by exonuclease degradation) until a region of homology is reached and recombination can occur. In an insertion vector with two regions of homology where one is short (< 500bp) then it can be envisaged that the longer region of homology is involved in pairing with the target site and initiating recombination, and the shorter region is then involved in the resolution of the event by enabling cross over to occur.

With a replacement vector, two crossover events mediated by two different regions of homology have to occur in order to introduce the positive selection marker at the site of integration, with concomitant loss of the negative selection marker (not shown). It is thought that the free end with homology to the genomic sequence will invade the target in the first cross over event, and a cross over between the second region of homology and the genomic target will complete the recombination event

## 1.6 Chromosome targeting vectors

### 1.6.1 Vector design

It was proposed that the L1 fragments chosen for incorporation into the vector were of sufficient length to direct HR, being 1.2kb + 1.4kb giving 2.6kb of total homology. The prototype vectors containing L1 fragments constructed by A.J. Brookes were of a standard PNS design, containing *neoR* for selection in mammalian cells but also containing *Leu2* for selection in yeast. A double cross over event would lead to removal of plasmid DNA and the negative selection marker HSV-tk (Wagner *et al.* 1981). The second generation of vectors was redesigned so that recombination through the repeat elements would lead to incorporation of several elements at the site of integration which would facilitate downstream manipulation and analysis. Integration of the bacterial-selectable marker *Amp* and the ColE1 origin of replication will enable rescue back into bacteria of a fragment surrounding the site of integration, and sites for unique cleavage of the genome will allow long range mapping of the region and the generation of specific fragments for recovery into bacteria or yeast (described in more detail below).

### 1.6.2 Alu repeats and recombination

It was questionable whether or not the Blur8 repeat would mediate HR. At 300bp it would perhaps not be long enough to direct HR according to the dsbr model, although it would be long enough to support a cross over event through a Holliday junction. Would the large number of potential targets increase the possibility of a HR event occurring or would the number of mismatches between the sequence and its potential targets counter this? There is evidence that Alu repeats are often involved in inter- and intra-chromosomal recombination events, as described below.

Alu repeat sequences have been found at the junctions of translocations, deletions and duplications of human chromosomes, which have been observed when the rearrangement has resulted in a disease phenotype and the chromosomes involved have been analysed in detail. Recombination between Alu repeats has been observed at loci with a high density of such repeats, for example, the low density lipoprotein (LDL) receptor protein gene (Lehrman *et al.* 1986, 1987a);  $\alpha$  galactosidase gene (Kornreich *et al.* 1990), C1 inhibitor locus (Stoppa-Lyonnet *et al.* 1990), and the apolipoprotein B gene (Huang *et al.* 1989). Most of the cases of recombination described have caused a deletion of sequences lying between two Alu repeats. Other Alu-Alu interactions have resulted in inversion-deletions (Li and Bray 1993), gene duplication (Lehrman *et al.* 1987a) and chromosome translocations (Onno *et al.* 1992). A single Alu repeat motif has also been found at recombination breakpoints involved in causing human disease for example in the LDL receptor gene (Lehrman *et al.* 1986) and in the  $\gamma$ -globin gene (Jagadeeswaran *et al.* 1992).

Although it is possible that a dsbr model of HR could be applied to these Alu/Alu interactions, it is unlikely considering that only 300bp of homology is involved. The observations of single Alu repeats at points of recombination suggests that Alu sequences may promote illegitimate recombination, i.e. are recombination hotspots. Breakpoints occurring within Alu elements tend to cluster within the LH repeat monomer (Lehrman *et al.* 1987b). This region contains the A and B boxes which promote transcription by RNA polymerase II. It has been suggested that this region may therefore constitute 'open' chromatin and be more prone to recombination (Lehrman *et al.* 1987a, Thomas and Rothstein 1989, Hu and Worton 1992, Nickloff 1992).

*Chi* sites, comprising the nucleotide octamer GCAGGAGG, are known to stimulate recombination of DNA within prokaryotes (Smith 1983). This octamer is a recognition site for the RecBC enzyme which introduces single strand nicks near the *chi* site as it unwinds DNA (Smith and Stahl 1985). A *chi*-like octamer has been associated with somatic recombination in human DNA, for example at rearrangements involving the oncogenes *bcl2* and *c-myc*, and also within the immunoglobulin heavy chain locus (Krowczynska *et al.* 1990). There is a motif with 7/8 bases identical to the *chi*-like octamer GC[A/T]GG[A/T]GG present within the Blur8 Alu repeat sequence, which also suggests that Alu repeat sequences may promote recombination for this reason alone.

There has been some work done by other groups to test the recombinogenic potential of Alu repeat elements. Wallenburg *et al.* (1987) transfected rat cells with a circular polyoma viral vector which contained both a repetitive rat sequence and a

unique mouse sequence. They found that integration of the plasmid preferentially occurred through the rat repetitive sequence, but that recombination was not directed into genomic repeat sequences and appeared to occur at random. Shen and Deininger (1992) devised a system in which the recombinogenic potential of two extra chromosomal sequences was tested. The sequences were inserted into two vectors containing different mutant tk genes which were then transfected into mammalian cells. They measured the numbers of colonies which survived HAT selection out of the total number of cells transfected, which had arisen as a consequence of recombination between the test sequences. Using this system they found that Blur8 and Blur11 sequences did not produce significantly more clones than did random control sequences and therefore concluded that Blur8 and Blur11 sequences do not contain hotspots for recombination in mammalian cells. This result has not to my knowledge been confirmed by other workers or by using alternative systems. Mechanisms for extra chromosomal recombination and for recombination into chromosomal sequences may be different, as discussed above.

Pavan and Reeves (1991) have used vectors containing the Blur8 repeat and a fragment from an L1 repeat to direct the integration of selectable markers into yeast artificial chromosomes (YACs) in *S. cerevisiae* cells. They found that both types of the human repeats could effectively direct recombination into human specific sequences. Using this method they were able to enrich 50 fold for yeast cells containing human sequences cloned into YACs, and also reduce the proportion of YACs containing rodent sequences which may arise when YAC libraries are created from somatic cell hybrids. This then provides evidence that despite its short length, Blur8 is able to direct species-specific recombination in yeast. In a similar strategy, Colleaux *et al.* (1993) have successfully used the mouse B2 repeat element to direct recombination of a plasmid containing a selectable marker and the recognition site for I-SceI into a YAC containing mouse sequences within a yeast cell. Targeted recombination could account for approximately 50% of the selected clones.

We wanted to test the ability of Alu repeat sequences to promote species-specific recombination in a somatic cell hybrid, whether it be through HR or through illegitimate recombination. The Blur8 element was included in three vectors: pTVBlur1 in which two repeat units were arranged in a head-to-head configuration; pTVBlur2, containing two repeat units in a head-to-tail orientation; and pLB in which a single Blur8 element was included in a vector also containing a single L1 fragment.



### 1.6.3 L1 repeats and recombination

There is no evidence *per se* for L1 repeats being involved in homologous recombination in mammalian cells. However the length of the 3' Line1 fragments in the vectors (1.8, 1.3 and 1.2kb) and the high number of very closely related target sequences (~2,000 copies with <5% variation from a consensus) fulfil the criteria for them to undergo homologous recombination in mammalian cells.

## 1.7 Vector features for manipulation of human genomic DNA

### 1.7.1 Sites for unique cleavage

#### *I-SceI*

I-SceI is a restriction endonuclease identified in *S. cerevisiae*. The endonuclease is encoded by a mobile group I intron from the yeast mitochondrial DNA which propagates itself by insertion into the double-stranded cut generated by the action of its endonuclease (Monteilhet *et al.* 1990). It has an asymmetric, 18bp recognition sequence. On a random DNA sequence with 50% GC content the probable occurrence of the I-SceI site is  $1.4 \times 10^{-11}$ , which is equivalent to approximately one site per DNA sequence 20 times the size of the human genome (Thierry and Dujon 1992). Therefore by chance we would not expect there to be an I-SceI site within the human DNA component of a monochromosome hybrid. The introduction of a site for I-SceI digestion into the human chromosome 11 by integration of our vector will therefore create a site at which the chromosome can be uniquely cleaved. I-SceI has previously been shown to uniquely digest human genomic DNA following introduction of its recognition site by both precise and random strategies (Kurdi-Haidar *et al.* (1993), Colleaux *et al.* 1993). It will therefore provide an anchor point for restriction maps of the region and allow the generation of sub-fragments of DNA with known ends which will facilitate cloning of the region.

#### *LacO (Achilles' Heel cleavage)*

The binding of the lac repressor protein to its recognition site, the lac operator (lacO) sequence in *E. coli*, has been well characterised. The lacO sequence can be modified to make it perfectly symmetrical which improves the affinity of the lac repressor protein for the site, and to also include a restriction site for the methylation-sensitive enzyme HhaI (Sadler *et al.* 1983). Koob *et al.* (1988) have shown that when bound to the modified lacO sequence, the lac repressor protein protects it from methylation by the modification methylase M.HhaI. When the methylase is removed and the repressor protein is uncoupled from its recognition site by the action of the inducer isopropyl  $\beta$ -D thiogalactoside (IPTG), the protected HhaI site is the only one which

remains unmethylated and therefore sensitive to cleavage by HhaI. This novel way of creating unique restriction sites at any position at which a lacO site has been introduced has been termed 'Achilles' heel' cleavage (Koob and Szybalski 1990, Grimes *et al.* 1990). It has been shown to be successful in cleaving yeast and *E.coli* genomes at a single site (Koob and Szybalski 1990, Garrison *et al.* 1992).

Hu and Davidson (1987) showed that the lac repressor would specifically bind the operator sequence when they were introduced into mammalian cells, and that the process could be reversed using the inducer IPTG. We therefore decided to include the lac operator sequence in the targeting vectors, as this would give an additional unique cleavage site in targeted clones, and it has the further attraction of offering the potential for affinity purification of the targeted locus, as described below.

### 1.7.3 Affinity Purification

The affinity for the lac repressor protein for its binding site may be exploited in another way to manipulate DNA after integration of the vector. Fusion proteins have been made which contain the recognition site of the lac repressor protein attached to the  $\beta$ -galactosidase ( $\beta$ -gal) protein, and are commercially available (Promega, Dynal). The  $\beta$ -gal part of the fusion protein can be recognised by specific antibodies which can then be detected by another set of antibodies which are conjugated to magnetic or immuno-beads. Levens and Howley (1985) utilised this to develop a method which allowed the isolation of DNA fragments containing a lacO sequence from a mixed population of DNA fragments in *E. coli*. The binding of the lac repressor/ $\beta$ -gal fusion protein to the lacO sequence allowed purification of the DNA fragment by immunoprecipitation. It is possible to apply this method to mammalian cells. Gossen *et al.* (1993) have enriched for a specific fragment containing a lacO site from restriction-digested mouse genomic DNA. The lac repressor/ $\beta$ -gal fusion protein was conjugated to magnetic beads and then added to the restriction-digested genomic DNA. After using a magnet to isolate the antibody/ fusion protein/ DNA complex, the purified fragments could be rescued into bacteria either by cloning into plasmids, or by simple circularisation and ligation of the fragment if it already contains the *Amp* and *ColE1* genes for bacterial selection (plasmid rescue). Gossen *et al.* reported an increase in efficiency of plasmid rescue of a specific fragment of 20 times that achieved without affinity purification.

It is hoped, then, that the introduction of a lacO site at the site of integration of the vector will facilitate rescue of specific fragments from around the region into bacteria. Selective YAC cloning of DNA from around the site of integration may also



be possible after enriching for the required fragments, and using the introduced *Leu2* gene as a marker selectable in yeast.

#### **1.7.4 *Leu2***

The *Leu2* gene incorporated into the targeting vectors was originally derived from *S. cerevisiae* (Andreadis *et al.* 1984). It has been shown to confer the ability to grow on leucine deficient medium on a *Leu2* auxotrophic yeast strain (Brookes *et al.* 1993), and therefore can be used as a marker selectable in yeast. This marker provides a means for selective recovery of a fragment of genomic DNA which encompasses an intact *Leu2* gene into a yeast artificial chromosome (YAC) (Burke *et al.* 1987). YACs allow detailed physical analysis of a large region of genomic DNA (Coulson *et al.* 1988), and they facilitate its manipulation when the ability of yeast cells to mediate efficient homologous recombination is utilised (Pavan *et al.* 1990). YACs can also be transformed into mammalian cells, and the genes transferred have been shown to be expressed (Gnirke *et al.* 1991, Gnirke and Huxley 1991, Huxley and Gnirke 1991). Complementation of a cellular defect, for example in a mouse fibroblast cell line deficient in collagen transcription, has been observed (Strauss and Jaenisch 1992). More recently phenotypic complementation in a whole mouse has been achieved after transfection of a YAC into mouse oocytes (Schedl *et al.* 1993).

Selective YAC cloning of a genomic fragment spanning an integration site of a targeting vector would therefore potentiate both physical and functional analysis of a region of interest.

## **CHAPTER 2**

## 2 Materials and Methods

### 2.1 Bacterial Culture

#### 2.1.1 Media and additives

All media was sterilised by autoclaving.

##### *L-Broth and agar*

2.46g MgSO<sub>4</sub>, 10g Bacto-tryptone (Difco), 5g yeast extract (Difco) and 10g NaCl were added per litre of distilled water. 15g Agar (Oxoid Ltd) was added per litre broth for L-agar.

##### *Terrific broth (TB).*

12g Bacto-tryptone, 24g yeast extract, 4g glycerol were added per 900ml of distilled water. After autoclaving, 100ml autoclaved phosphate solution (0.1M potassium dihydrogen phosphate and 0.72M dipotassium hydrogen phosphate) were added.

##### *Xb medium*

5g yeast extract (DIFCO), 20g bacto-tryptone (DIFCO), 5g magnesium sulphate septahydrate, 0.76g potassium chloride were added per litre of distilled water. 14g per litre of Bacto agar was added for Xa medium

##### *Ampicillin*

Ampicillin was added to agar and broth in order to select for bacteria carrying resistance to this antibiotic conferred by a plasmid. A stock solution of ampicillin (Sigma) was made up at a concentration of 50mg/ml. This solution was filter sterilised and stored at -20°C. It was added to broth and agar to give a final concentration of 50µg/ml.

##### *5-Bromo-4-Chloro-3-Indolyl β -D-Galactopyranoside (X-gal)*

X-gal acts as a substrate for β-galactosidase. It was added to agar for the growth of plasmids carrying blue/white colour selection. A Stock solution of X-gal was made up at a concentration of 20mg/ml mg/ml in DMSO and stored, protected from light at -20°C. It was added to agar to give a final concentration of 40 µg/ml.

##### *Isopropyl β -D-Thiogalactopyranoside (IPTG)*

IPTG is a derepressor of the Lac operon. It was added to agar for the growth of plasmids carrying blue/white colour selection. 1.2µl of a 100mM stock solution of IPTG (Sigma) was added for every ml of L-agar required.

#### 2.1.2 Bacterial Strains

**DH10B** genotype: F<sup>-</sup>, araD139, Δ(ara, leu)7697, ΔlacX74, galU, galK, mcrA Δ(mrr-hsdRMS- mcrBC), rpsL, dor, φ80dlacZΔM15, endA1, nupG, recA1.

This strain was derived from MC1061 (as described by Grant *et al.* 1990) and is mutated in loci encoding cytosine and adenine methylation dependent restriction

systems (MDRS). It has been shown to give higher frequencies of plasmid rescue from mammalian DNA (Grant *et al.* 1990), and was used as a host for some pBS and pBSK vectors as well as in all plasmid rescue experiments.

**DH5 $\alpha$**  genotype: F<sup>-</sup>endA1, hsdR17(r<sub>k</sub><sup>-</sup>,m<sub>k</sub><sup>-</sup>), supE44, thi-1 $\lambda$ <sup>-</sup>,recA1, gyrA96, relA1, deoR,  $\Delta$ (lacZYA-argF)-U169,  $\phi$ 80dlacZ $\Delta$ M15.

This strain was derived from DH5 (as described by Woodcock *et al.* 1989), and was used as the host strain for some of the vector constructs.

**JM83** F<sup>-</sup>, ara  $\Delta$ (lac-proAB), rpsL, (Str<sup>r</sup>), ( $\phi$ 80d $\Delta$ (lacZ)M15)

This was the host strain for some of the vector constructs (described by Vierra and Messing 1982).

### 2.1.3 Plasmid vectors

Bluescribe (pBS, Stratagene) and Bluescript SK (Stratagene) are plasmid vectors based on pUC19 (Vierra and Messing 1982). They contain the bacterial origin of replication (ColE1) which allows replication in bacteria to high copy number, and the ampicillin resistance gene (Amp<sup>R</sup>) which enables selection for plasmid transformation. The vectors carry two different polylinkers which are located within a  $\beta$ -galactosidase (lac Z) gene in the plasmid in frame. Disruption of lac Z by cloning into the polylinker results in white colonies being produced upon induction by IPTG and growth on the chromogenic substrate Xgal. Plasmids containing an insertion therefore produce colonies which can be differentiated from blue wild type colonies.

## 2.2 Transformation of bacteria

### 2.2.1 Preparation of cells for heat shock transformation

This method was a modified version of that described by Hanahan (1983).

A sterile loop was used to streak out a bacterial host strain from a frozen glycerol stock onto an Xb-agar plate. The following day a single largish colony was used to inoculate 5 ml of Xb-broth in a Universal tube. This was incubated overnight at 37°C shaking at 250rpm. The next morning 2ml of the overnight culture was used to inoculate 100ml pre-warmed Xb broth (a ratio of 1:50 ) in a 500ml flask. This culture was then re-incubated at 37°C with shaking until it attained an OD<sub>550</sub> of 0.6 (between 60-120 min depending upon the host strain being used). At this point the culture was transferred to a pre-chilled sterile 500ml flask and incubated on ice for 10-20 min with gentle swirling. This was to ensure rapid chilling of the cells which halts cell division. The suspension was split between pre-chilled 50ml Falcon tubes and the cells harvested by spinning for 5 min at 3000rpm at 4°C. The supernatant was carefully removed, and each pellet resuspended in 20ml ice-cold buffer I (30mM

potassium acetate, 100mM rubidium chloride, 10mM calcium chloride, 50mM manganese chloride, 15% glycerol (v/v), pH to 5.8 with acetic acid filter-sterilised and stored at 4°C), using a pre-cooled sterile pastette. The suspension was incubated on ice for 5 min, before harvesting the cells again by spinning in a centrifuge for 5 min at 2500 rpm at 4°C. The supernatant was carefully removed and each cell pellet gently resuspended to a final volume of 2 ml of buffer II (10mM PIPES, 75mM calcium chloride, 10mM rubidium chloride, 15% glycerol (v/v), pH to 7.0 with potassium hydroxide, filter-sterilised and stored at 4°C). The suspension was incubated on ice for 15 min after which small volumes (50-100µl) were aliquoted into Eppendorf tubes sitting in a dry ice/ethanol bath so that the cells were snap-frozen. The cells were then stored at -70°C.

### **2.2.2 Heat shock transformation of bacteria**

An aliquot of frozen competent cells was thawed on ice. The plasmid ligation in a volume of less than 20µl was added to the cells and mixed gently. The cells were then incubated on ice for about 20 min, and then subjected to heat shock by incubating them at 42°C for 90 seconds. They were then immediately chilled on ice for a further 2 min. L Broth was then added to the cells to a final volume of 1ml before they were incubated at 37°C with constant shaking (250rpm). 100µl of the cell suspension was streaked out per L agar plate with the appropriate antibiotic selection. The plates were incubated at 37°C overnight. Typically 10<sup>5</sup> colonies per ng of covalently closed circular (ccc) vector were obtained by this method.

### **2.2.3 Preparation of competent cells for electro-transformation**

An overnight culture of the appropriate bacterial strain was prepared as for heat-shock competent cells except 5ml of terrific broth with appropriate selection was inoculated. 200ml of terrific broth pre-warmed in a 500ml flask was inoculated with 2ml of the overnight culture (1:100). The cells were grown to an OD  $\lambda_{550}$  of 0.6-0.8, depending on the strain being used. For DH10B OD  $\lambda_{550}$  = 0.72. The culture was transferred to a 500ml flask which had been previously chilled on ice. and the culture was then incubated on ice for 10 min. The suspension was divided between 4 chilled Falcon tubes and centrifuged at 4°C for 15 min at 3500 rpm in a Sorvall bench top centrifuge to harvest the cells. Each pellet was resuspended in 50 ml of water at 4°C. This water had been de-ionised, UV-sterilised, and autoclaved in bottles with no trace of detergent then stored at 4°C. It is very important that the water does not contain any impurities which will alter its pH from 7.0 or any detergents. The suspension was then respun as before and the supernatant carefully removed. Each pellet was



resuspended in 25ml of the sterile water, containing 10% glycerol. The cells were again pelleted as before. The supernatant was carefully removed to leave the pellet in as small a volume of buffer as possible. The pellets were pooled and resuspended to a final volume of approximately 800 $\mu$ l. 50  $\mu$ l aliquots of the suspension were dispensed into Eppendorf tubes sitting in a dry ice/ethanol icebath. The cells were stored at -70°C. Following this protocol transformation frequencies of  $1 \times 10^8$  per ng ccc DNA were often obtained and in some cases this rose to  $5 \times 10^8$ .

#### **2.2.4 Transformation of bacterial cells by electroporation**

An aliquot of prepared frozen cells was thawed on ice. The ligation mix to be transformed into the cells was dialysed against 1x TE using a 0.005 $\mu$ m Millipore filter (VS). 1-2 $\mu$ l of the prepared ligation was gently mixed into the thawed cell suspension and incubated on ice for approximately 1min. The mixture was transferred to a chilled, sterile 0.2cm electroporation cuvette which was then placed in a Gene Pulser apparatus (BioRad), previously set up according to the manufacturers instructions. A field strength of 12.5 kV/cm was applied across the cells after which 1ml of L broth was immediately added. They were then incubated at 37°C for 1 hr with constant shaking at 250rpm. 100 $\mu$ l of the suspension was streaked out on L agar plates with the appropriate antibiotic selection and incubated at 37°C overnight.

### **2.3 Preparation of Plasmid DNA**

#### **2.3.1 Small-scale plasmid DNA preparation**

This is an adaptation of a small-scale alkaline-lysis method of plasmid preparation described by Jones and Schofield (1990).

A single bacterial colony was picked into 5ml of Terrific Broth (TB) with appropriate antibiotic selection and incubated overnight at 37°C with vigorous shaking. 1.5ml of the culture was decanted into an Eppendorf tube and the cells pelleted by centrifugation at 13000 rpm for 2 min. The supernatant was decanted from the pellet and the tube allowed to drain for several min to remove any remaining fluid. The pellet was resuspended in 150 $\mu$ l of Solution I (50 mM glucose, 10mM EDTA, 25mM Tris pH 8.0, filter sterilised) by vortexing. 300 $\mu$ l of Solution II (0.2M sodium hydroxide, 1% SDS made fresh just before use) was then added and the tube inverted several times to mix the contents before placing it on ice for 5 min. 225 $\mu$ l of Solution III (5 M acetate, 3M potassium ions, sterilised and stored at 4°C) was added and mixed again by inversion of the tube before replacing it on ice for a further 5 minutes. The tube was then centrifuged at 13000 for 5 min. One volume of absolute ethanol was added to the supernatant decanted into a fresh tube. Immediately this tube was



again centrifuged at 13000 for 5 min at 4°C. The pellet thus obtained was drained and washed in 70 % ethanol before being dried under vacuum for 5 minutes. The pellet was finally resuspended in 40µl of 1x TE.

DNA obtained in this way could be used in restriction digest analysis or directly in a sequencing reaction (see 2.12.2). The average yield was between 20 and 50µg of ccc plasmid DNA per miniprep.

### **2.3.2 Midi-scale plasmid DNA preparation**

This method was performed as described by Noguchi (1990).

A single plasmid colony was used to inoculate 50ml terrific broth containing ampicillin in a 200ml conical flask. The flask was incubated overnight at 37°C with constant shaking at 250rpm. The culture was spun down in a 50 ml Falcon tube for 10 minutes at 3000rpm at 4°C. The pellet was retained and rinsed by resuspending in 2ml ice cold 1xTE buffer split between two Eppendorf tubes. The cells were then repelleted by centrifugation for 1 min at 13000 rpm and each pellet resuspended in 300µl ice cold STE (15% sucrose, 50mM Tris HCl pH 8.0, 1mM EDTA pH8.0, filter-sterilised and stored at 4°C) with 100µl lysozyme (5mg/ml in STE). The tubes were then incubated on ice for 10 min, inverting them occasionally to mix the contents. 370µl ice cold TTE (50mM Tris HCl pH 8.0, 50mM EDTA pH8.0, 0.1% Triton X-100, filter sterilised and stored at 4°C) was then added to each tube which were then incubated for a further 10 min, again inverting the tubes occasionally. 2µl DEPC (diethylpyrocarbonate) was added to each tube; this denatures proteins present. The tops of the tubes were punctured before they were then immediately incubated in a boiling water bath for 45s. Straight from the water bath, the tubes were spun at 13000rpm for 15 min at room temperature in a benchtop Eppendorf-centrifuge. The soft bulky pellet was removed using a sterile toothpick and was discarded. 1ml of absolute ethanol was added to the remaining supernatant, and immediately after mixing by inversion, the tubes were spun for 3 min at room temperature. The supernatant was discarded and the pellets were allowed to air dry. The pellets were then pooled in 700µl distilled water.

After the pellet had resuspended, 5µl of 10mg/ml RNaseA was added to the suspension, and it was then incubated at 37°C for 30 min. 1µl of 10mg/ml Proteinase K was added to the tube which was then incubated at 50°C for a further 15 min. 35µl 5M sodium chloride was added to the suspension to bring it to a final concentration of 2M to facilitate phenol extraction. The solution was then extracted once with phenol by mixing it with an equal volume of phenol (saturated in 1x TE pH 8.0) and retaining the upper, aqueous layer after a brief centrifugation. The mixture was then extracted

twice with a mixture of phenol: chloroform: isoamylalcohol (50:49:1), again added in an equal volume, mixed and the upper aqueous layer retained after a brief centrifugation. Finally it was extracted once with chloroform: isoamylalcohol (49:1), retaining the upper aqueous layer after mixing and centrifugation. DNA was then precipitated from the solution by adding 800µl ice-cold absolute ethanol and spinning the tube immediately at 13000rpm for 5 min at room temperature. The pellet was then rinsed in 70% ethanol, vacuum dried and finally resuspended in an appropriate volume of 1X TE. We routinely obtained around 300µg of covalently closed circular (ccc) plasmid DNA from a 50 ml culture.

### **2.3.3 Large scale preparation of plasmid DNA**

This was from a method described by Maniatis *et al.* (1989).

400ml of terrific broth in a 1 litre flask containing the appropriate antibiotic selection (50µg/ml ampicillin) was inoculated with a single large colony of the host strain containing the plasmid required. This was incubated overnight at 37°C with continuous shaking at 250rpm. Cells were harvested in 2 x 200ml clean Sorvall bottles, by centrifugation at 6000rpm for 10 min at room temperature. The supernatant was decanted and each pellet was resuspended in 9ml Solution I (50mM glucose, 25mM Tris pH 8.0, 10mM EDTA pH 8.0 filter-sterilised). To this, 1ml of lysozyme at 10mg/ml made up freshly in 10mM Tris pH 8.0 was added, and the suspension was mixed and incubated at room temperature for 10 min to initiate cell lysis. 20ml of Solution II (0.2M sodium hydroxide, 1% SDS freshly made up) was then added to each bottle and was swirled to mix before incubating on ice for 5 min to complete lysis of the cells. 15ml of ice-cold solution III (5M acetate, 3M potassium, sterilised by autoclaving) was then added to the bottles to precipitate the unwanted bacterial chromosomal DNA. After shaking vigorously to ensure mixing, the solution was then incubated on ice for a further 15 min. The cell lysate was centrifuged at 4000rpm for 15 min at 4°C; the rotor was allowed to stop without the brake to minimise dislodging of the pellets. The supernatant was filtered through muslin into a clean 200ml centrifuge bottle. 0.6 volumes of isopropanol was added to the supernatant and mixed before incubating at room temperature for 10 min. The precipitated plasmid DNA was then recovered by centrifugation at 5000rpm for 15 min at room temperature. The pellet was drained and rinsed in 70% ethanol before drying under vacuum.

The pellet was resuspended to a final total volume of 7ml in 1 x TE (10mM Tris HCl pH 7.5, 1mM EDTA). 7g of caesium chloride was dissolved in this solution and 100µl of ethidium bromide (10mg/ml) was added, to give a final refractive index

of approximately 1.55g/ml. This solution was divided between two quickseal ultracentrifuge tubes. The tubes were centrifuged in a fixed-angle rotor for 16-18 hr at 100,000rpm. The ccc plasmid DNA should be visible as a tight band about half way up the centrifuge tube, without the aid of a UV light source. The band was extracted from each tube using a sterile 0.8 gauge hypodermic needle attached to a syringe and pooled in a polypropylene tube. The ethidium bromide was removed by repeated extraction with water-saturated butan-1-ol, discarding the organic phase until no pink colour remained. The DNA was precipitated from the caesium chloride-containing solution by diluting it with 3 volumes of sterilised water and then adding 2 volumes of ethanol. The solution was mixed and left at 4°C for 15 min before centrifuging for 30 min at 4°C and 10,000g. The pellet of DNA was drained and vacuum-dried before resuspending it in 500µl of 1 X TE. The  $OD_{\lambda_{260}}$  was measured to estimate the concentration of plasmid DNA present. Routinely 1.5-2.5 mg of ccc plasmid DNA was obtained.

## 2.4 Enzymatic manipulation of plasmid DNA

### 2.4.1 Restriction endonuclease digestion of DNA

Reactions using restriction enzymes were carried out in KGB buffer (100mM potassium glutamate, 25mM tris-acetate pH 7.6, 10mM magnesium acetate, 50µg/ml BSA, 0.5mM 2-mercaptoethanol, filter sterilised and stored at 4°C, described by McClelland *et al.* 1988), when ever possible. In other cases, digests were carried out in the appropriate buffer according to the manufacturers instructions. Where two different enzymes were to be used, digests were carried out simultaneously if both enzymes required the same buffer. Otherwise, digestion with the enzyme requiring the lower salt concentration in its buffer was carried out first, then the salt concentration altered by adding the appropriate amount of sodium chloride solution. Reactions were terminated by heating the reaction to 68°C or 80°C for 15 min according to the sensitivity of the enzyme (NEB catalogue 1992) or by phenol extraction.

### 2.4.2 Dephosphorylation of linear plasmid DNA

7µg of linear plasmid DNA in a volume of 45µl water was dephosphorylated with 0.1 units of calf intestinal phosphatase (CIP, B.M.) and 5µl CIP buffer (10x is 1mM EDTA, 0.5M Tris-HCl, pH 8.5 supplied by B.M.). The reaction was incubated at 56°C for 30 min then another 0.1 unit CIP was added and incubation continued for a further 30 min. The reaction was stopped by the addition of 1ml 0.5M EDTA (to a final concentration of 10mM), (Cobianchi and Wilson 1987). The volume was brought up to 150µl by the addition of 1x TE before performing 2 phenol:

chloroform: isoamylalcohol extractions, followed by a single chloroform: isoamylalcohol extraction. The DNA was then precipitated by adding an equal volume of ice-cold absolute ethanol and 1/10th volume 3M sodium acetate, chilling at -20°C for 30 min then centrifuging at 15000rpm for 15 min at 4°C. The pellet was washed in 70% ethanol and then resuspended in 10ml 1x TE.

#### **2.4.3 Preparation of plasmid insert DNA**

Typically, 5µg of plasmid DNA was fully digested with the appropriate restriction enzyme. The insert was isolated by running out on an agarose gel and purified as described in section 2.7.3.

#### **2.4.4 Annealing of oligonucleotides**

1µg of oligo 1 and 1 µg of oligo 2 were combined with 1µl 10x TM buffer (10X is 100mM Tris, 10mM magnesium chloride ) in a total of 10ml of water. The reaction was heated to 68°C for 2 min then allowed to cool slowly in a beaker of water until room temperature was reached. Annealing was checked by running an aliquot on a 4% nuseive agarose gel. The annealed oligos were stored at -20°C.

#### **2.4.5 Conversion of 5' overhangs to blunt-ended molecules**

Bacteriophage T4 DNA polymerase has a 5' to 3' polymerase activity, and a 3' to 5' exonuclease activity. This is utilised to 'fill in' recessed 3' ends of DNA molecules to generate blunt-ended molecules, (Cobianchi and Wilson 1987), alternatively the Klenow fragment of *E.coli* DNA polymerase I can be used as it has a similar activity. Approximately 5µg of plasmid DNA, linearised with a restriction enzyme which has generated a 5' overhang was diluted in a total of 28.5 µl water. To this was added 1.5 µl 50 x GATC mix (dGTP, dCTP, dATP, dTTP each to a final concentration of 400µM) and 3 units Klenow fragment (B.M., sequencing grade). In other cases, ~750ng of annealed oligonucleotides was diluted in 5.5µl of water, with 2µl 5x GATC mix 1µl 10x buffer (B.M. buffer 'A') and 1.5 units T4 DNA polymerase. With either set of substrates, the reaction was incubated at 37°C for 60 min, after which the enzyme was inactivated by incubation at 68°C for 20 min.

#### **2.4.6 Ligation of sticky-ended molecules**

The optimal ratio of ends for such a reaction is 1 : 2 of vector : insert (Cobianchi and Wilson 1987). Typically 160ng dephosphoylated, linearised vector (@ 4kb) was ligated to 80ng insert fragment (@ 1kb) with 1µl 10 x ligase buffer (1 x buffer is 50mM Tris pH 7.4, 10mM magnesium chloride, 10mM DTT, 1mM spermidine, 1mM

ATP 100µg/ml BSA) and 0.1 unit T4 DNA ligase (B.M.) in a total reaction volume of 10µl. The reaction was incubated overnight at 16°C.

#### **2.4.7 Ligation of linkers (sticky-ended)**

The reaction conditions were essentially the same as for other sticky-ended ligations, but in this case the linkers or annealed oligos are usually of the order of 20bp so the relative amount of vector has to be increased (Cobianchi and Wilson 1987). Typically, 800ng of non-dephosphorylated, linearised vector was combined with 8ng of annealed oligos in the same reaction mix as above. If phosphorylated linkers were being used then the linearised vector was dephosphorylated prior to the ligation. The reaction was incubated for a minimum of 3 hours at 16°C.

#### **2.4.8 Ligation of blunt-ended molecules**

The optimal ratio of ends for such a ligation event is 3 : 1 of vector : insert, (Cobianchi and Wilson 1987). Typically, 160ng dephosphorylated, linearised vector (@ 4kb) was combined with 13ng insert (@ 1kb). 1µl 10 x T4 DNA ligase buffer (Maniatis et al.1989, 1x buffer is 20mM Tris pH 7.6, 5 mM magnesium chloride, 5 mM DTT, 50 mg/ml BSA), 30mM potassium chloride, 10mM ATP and water to 10µl and 0.1 unit T4 DNA ligase were added and the reaction was incubated overnight at 4°C.

#### **2.4.9 Ligation of linkers (blunt-ended)**

The reaction conditions were similar to those used in other blunt end ligations except that PEG 8000 was included to increase macromolecular crowding to improve the efficiency of the reaction. Typically, 100ng linearised vector was ligated to 1ng linker DNA with 2µl 10x ligase buffer (Maniatis *et al.* 1989), 10% PEG 8000, 0.5mMATP, 30mM potassium chloride, 0.1 unit T4 DNA ligase, and water to 20µl. The reaction was incubated overnight at 20°C.

#### **2.4.10 Digestion with I-SceI**

I-SceI (Boehringer Mannheim) is a restriction enzyme which requires magnesium chloride in order to act on its substrate, but which is at the same time unstable in the presence of magnesium ions. Plasmid DNA to be digested with the enzyme was therefore equilibrated in 1X buffer (100mM Diethanolamine/ HCl, 2mM 1,4 dithiothreitol (DTT), 0.2 mg/ml BSA, 2mM spermidine/HCl, pH 9.5) and 5 units/µg DNA of I-SceI enzyme and the reaction mixed before adding magnesium chloride to a final concentration of 5mM. The reaction was then incubated for 45 min at 37°C,



after which the reaction should be complete and the activity of the enzyme will be reduced.

## **2.5 Preparation of vector DNA for transfection**

### **2.5.1 Restriction digestion**

Approximately 500 $\mu$ g of DNA prepared by maxiprep was linearised by restriction digest with the appropriate enzyme. Completion of the reaction was confirmed by comparing the electrophoretic mobility of an aliquot of the treated DNA to that of an uncut sample.

### **2.5.2 Bal 31 exonuclease digestion**

Typically, 200 $\mu$ g of linear DNA was suspended in Bal31 buffer (1 x is 20mM Tris HCl, 600mM sodium chloride, 12.5mM magnesium chloride, 12.5 mM calcium chloride, and 1mM EDTA, pH 7.2), in a total volume of 400 $\mu$ l with 6 units of Bal 31 exonuclease (B.M.). Aliquots of 50 $\mu$ l or 100 $\mu$ l were removed into 20 $\mu$ l 0.5M EDTA at 4°C at various intervals after starting the reaction, to terminate it. The aliquots were then diluted with water to give a final concentration of 5mM EDTA. The extent of the exonuclease treatment was determined by running approximately 800ng of each aliquot out on an agarose gel, transferring the DNA to a nylon membrane by Southern blot and then hybridising the blot to an oligonucleotide probe (C308) complementary to the polylinker sequence at the end of the linear vectors. The length of incubation required to just remove the linker sequence from most of the molecules was then further defined, by repeating exonuclease digestion and removing aliquots at shorter time intervals within the range defined by the first experiment. The required aliquots of exonuclease-treated DNA were pooled giving approximately 20 $\mu$ g of DNA per  $2.5 \times 10^6$  cells to be transfected.

### **2.5.3 Purification and concentration of Bal31-treated DNA**

To concentrate the DNA and remove unwanted salts, the solution was passed through a Centricon-30 column (Amicon), following the manufacturers instructions. The column was centrifuged at 6500rpm in a Sorvall SS34 rotor for 6 min to wash the sample, and then at 2500rpm for 2 min to obtain the concentrated sample from the column.

Alternatively, salt was removed from the samples by passing them through a 'Magiprep' column (Promega), using a 1ml syringe to force the solution through the column (see manufacturers instructions).



In a more simple strategy, DNA was cleaned by extraction with phenol/chloroform, chloroform, and ether, followed by precipitation with 100% ethanol and rinsing of the pellet with 70% ethanol to remove unwanted salts before re-suspension of the DNA.

## **2.6 Mammalian Cell Culture**

### **2.6.1 Cell lines**

J1 clone 4 (J1c4) is a human-Chinese hamster ovary somatic cell hybrid line containing a single human chromosome 11 as the sole human component (Kao *et al.* 1976).

GM10321 and GM10498 are human-mouse fibroblast hybrid cell lines, each containing a single human chromosome 17 (provided by John Boyle).

### **2.6.2 General Maintenance**

J1c4 was grown in RPMI 1629 (McCoy's 5a medium) supplemented with 10% foetal calf serum (FCS) at 37°C under 5% CO<sub>2</sub>. GM10321 and GM10498 were grown in Dulbecco's modified minimal essential medium (DMEM) supplemented with non-essential amino acids, 1mM sodium pyruvate, 4mM glutamine, 10% FCS and HAT (1x10<sup>-4</sup>M Hypoxanthine, 1x10<sup>-7</sup>M methotrexate, 1.6x10<sup>-5</sup>M thymidine). Cells were cultured at 37°C under 10% CO<sub>2</sub>. Sterile culture flasks, Universal tubes, Nunc freezing vials and glassware were used throughout. All work was carried out under a lamina flow hood. All cells were harvested from flasks by rinsing the monolayer with a minimal amount of Dulbecco's phosphate buffered saline (PBS), then incubating the cells with enough 10% trypsin in versene solution to just cover the monolayer at 37°C to facilitate loosening of the cells. Cells were harvested from suspension by centrifugation at 1000rpm for 5 min, and resuspended in the appropriate amount of culture medium.

Cells which had been previously frozen down in vials and stored under liquid nitrogen (-70° C) were rapidly thawed by placing the vial in a beaker of water previously warmed to 37°C. When the cell suspension had completely thawed, the storage medium (foetal calf serum: FCS, 10% dimethylsulphoxide :DMSO) was diluted with 5 ml culture medium prior to gentle centrifugation. Cells were initially plated into a 25 ml dish with 10ml medium.

To store cells at -70°C, cells were harvested and resuspended in 2 ml of freezing mix, previously chilled on ice. Approximately 1ml of cell suspension from one 80% confluent 25cm<sup>2</sup> flask was aliquoted into each of 2 Nunc freezing vials also previously chilled on ice. The cells were initially frozen to -20°C in a padded box to ensure even

slow freezing and avoid cell damage from water crystallisation. After at least 24 hr at -20°C they were transferred to liquid nitrogen for long-term storage.

### **2.6.3 Transformation of cells by lipofection**

Cells to be transformed were grown to 50-60% confluence in a 25cm<sup>2</sup> or 75cm<sup>2</sup> flask. Volumes of reagents in smaller scale transfections were approximately 1/3 those used in the larger scale experiments. For most transformations, a 75cm<sup>2</sup> flask was used for which approximately 30µg of prepared DNA per flask was brought to a volume of 150µl with distilled water. The cationic lipid DOTMA (N[1-(2,3-dioleyloxy)propyl]-N,N,N-trimethylammonium chloride) can form liposomes, and can interact spontaneously with DNA to form a liposome/polynucleotide complex (Felger *et al.* 1987). The resulting complexes can deliver functional nucleic acid molecules to cells in tissue culture. DOTMA is marketed by BRL under the name of Lipofectin. According to the manufacturer's instructions, 90µg of Lipofectin™ BRL, (1mg/ml), was diluted with 60µl distilled water. The two reagents were then mixed together gently, in a polystyrene bijou bottle and allowed to stand at room temperature for 15 min, to promote the formation of lipid-DNA complexes. The cells were washed twice in RPMI or DMEM to remove traces of FCS which would otherwise inhibit the transfection procedure. Only 3ml of serum-deficient medium was finally aliquotted into the flask after which the lipofectin-DNA complex mixture was added dropwise, whilst gently swirling the flask. The cell monolayer was incubated at 37°C for between 6 to 9 hours to allow the uptake of DNA. After this time, FCS and RPMI was added to a final volume of 20 ml with a final concentration of 10% FCS. 24 hr later this medium was replaced with selection medium containing RPMI, 10% FCS, 0.4mg/ml G418 (GIBCO), 6µM gancyclovir (Syntex Pharmaceuticals Ltd.). The medium was replaced every 2-3 days to prevent the accumulation of dead cells.

### **2.6.4 Selection conditions**

Two exogenous selectable markers were used for the positive and negative selection of transfected cultured cells. The bacterial gene *neoR*), confers resistance to the aminoglycoside antibiotic G418 in mammalian cells (see section 1.1.2) and was the positive selectable marker in our vectors. Kill curves were performed on untransformed cell lines to establish the minimum effective concentration of the particular batch of G418 used. This was in the range of 0.4mg/ml. Gancyclovir is a base analogue which is toxic to cells upon metabolism by the product of the herpes simplex thymidine kinase (HSV-tk) gene, but it is not a substrate for the endogenous mammalian thymidine kinase gene. It was used as a means of selection against clones

containing the HSV-tk gene. The maximum concentration tolerated by untransformed cells was established. This was in the range of 6 $\mu$ M.

### **2.6.5 Picking resistant clones**

After 2-3 weeks of incubation under the established conditions, foci of cells comprising of a colony which had arisen from a single resistant cell were visible. When these had reached 1-5mm in diameter they were ready to be picked. The medium was replaced with a minimal amount of PBS, just sufficient to cover the cells. The top of the flask was removed using a hot-blade apparatus. A colony was gently scraped together using a sterile, fine-tipped plastic pastette and then sucked up into its tip. The cells were deposited into 100 $\mu$ l trypsin/versene. After a few minutes they were transferred to a 2cm well containing 3ml of selection medium. As many as possible of the clones were harvested in this way, the remainder were harvested and frozen down together using trypsin and versene as previously described. Clones were expanded to a 25cm<sup>2</sup> flask when they reached confluence which took between 1-3 weeks.

## **2.7 Manipulation of Genomic DNA**

### **2.7.1 Rapid isolation of DNA from mammalian cells**

This protocol was taken from Laird *et al.* (1991). Cells were harvested from a 75cm<sup>2</sup> flask and stored as a pellet at -20°C until required in a 10 ml conical-bottomed Sterilin tube. The cell pellet was thawed then gently resuspended in 300-500 $\mu$ l lysis buffer (100mM Tris-HCl pH 8.5, 5mM EDTA, 0.2% SDS, 200mM sodium chloride, 100 $\mu$ g/ml proteinase K). The suspension was incubated at 37°C for several hours or overnight. An equal volume of isopropanol was added to the lysed cells and the tube was gently mixed until the interface disappeared. The DNA was spooled out using a glass rod and after rinsing in 70 % ethanol was allowed to air dry. The DNA was resuspended in 300-500 $\mu$ l 1 x TE. Incubation at 55°C for 2 hours facilitated resuspension of the DNA.

### **2.7.2 Isolation of high quality DNA from mammalian cells**

Cells were harvested as described in 2.7.1. The cell pellet was resuspended in a small volume (100-200 $\mu$ l) of cell suspension buffer (100mM Tris, 100mM EDTA, 150mM sodium chloride). Lysis buffer (1% SDS, 10mM Tris pH 8.0, 150mM sodium chloride, 10mM EDTA) was added to a total of 10x the original volume i.e. 1-2ml. Ten microlitres of RNase (10mg/ml) was added for every ml of cell suspension, and

the tube was incubated at 37°C for 15 min. Proteinase K was then added to the solution to a concentration of 2mg/ml and the solution was then incubated at 50°C for 3hr. After incubation, an equal volume of phenol (saturated with 1x TE pH 8.0) was added and mixed with the solution to give a milky emulsion. The tube was centrifuged at 2500rpm for 10 min to separate the phenol containing extracted proteins from the solution containing the DNA. The upper aqueous layer was removed into a clean tube and extracted with an equal volume of phenol: chloroform: isoamylalcohol (50: 49: 1) by mixing and centrifuging as before. The upper aqueous layer was again transferred to a clean tube before the final extraction of contaminating proteins was made with an equal volume of chloroform: isoamylalcohol (49: 1). DNA was precipitated from the aqueous layer which contained contaminating salts by transferring the solution to a new tube and adding a tenth volume of sodium acetate with 3 volumes of absolute ethanol. After gently mixing the solutions, DNA was spooled out on to a clean glass rod. It was then cleaned by rinsing in a solution of water: ammonium acetate: ethanol (1: 0.5: 3) and allowed to air dry on the rod before transferring to ~500µl 1x TE. The DNA was allowed to resuspend by incubation at 4°C with gentle agitation.

### **2.7.3 Restriction endonuclease digestion of genomic DNA**

Typically, 10 µg of genomic DNA was digested in not less than four volumes of buffer; usually between 30-50µl. In a 40µl reaction 10 µg genomic DNA was combined with 4µl 10 x appropriate restriction enzyme buffer, 4 µl bovine serum albumen (BSA) Fraction V (B.M.) (1mg/ml), and 15 units of the required restriction enzyme. 2µl spermidine (100mM) was also added to reactions carried out in buffer containing >50mM sodium chloride. The final volume was made up to 40µl with water. The reaction was then incubated for 2 hr under the conditions specified by the manufacturer for the enzyme, after which 5µl loading buffer (20% ficoll, 100mM EDTA, orange G) was added to stop the reaction. The reactions were stored at -20°C until required.

Restriction digests with I-SceI were carried out slightly differently. 10µg of DNA was suspended in 1x I-SceI buffer (see section 2.4.10) and mixed with ~25 units of I-SceI restriction enzyme in a final volume of 30-50µl. 3-5µl 50mM magnesium chloride was then added to give a final concentration of 5mM. The reaction was incubated for just 45 min at 37°C. If double digests were to be performed, DNA was always digested with I-SceI first before precipitating the sample and resuspending it in a buffer appropriate to the enzyme for the second reaction.



#### 2.7.4 Preparation of chromosomal DNA in agarose microbeads

This was from a method described by Cook (1984).

Cells were grown until they had just reached confluence in 225cm<sup>2</sup> flasks. Typically 4 flasks were used in the production of a single batch of microbeads. The cells were harvested into Falcon tubes and the pellets were washed twice in PBS, resuspending the cells each time in fresh PBS to a concentration of approximately  $5 \times 10^7$ . The cells were then pooled into one tube and the total number was estimated using a haemocytometer. They were finally resuspended to a concentration of around  $1.25 \times 10^7$  cells/ml. The cell suspension was warmed to 37°C, and three drops of polybead-dyed polystyrene microspheres (Polysciences) were added per 5ml of suspension. One volume of cell suspension was added to 0.25 volumes of 2.5% LMT ultra pure agarose (BRL) (previously cooled to 50°C) and 2.5 volumes of liquid paraffin (previously warmed to 50°C). The solutions were combined in a round-bottomed flask (warmed to 37°C) and the flask and contents were then immediately agitated on a flask shaker for 30s at full speed. The flask was then transferred to an ice/water bath and continuously swirled for 10 minutes. The contents were poured into 50ml falcon tubes, collecting any residual solution by rinsing out the flask with PBS. The tubes were topped up with PBS and spun in a Sorvall RT6000 at 3000rpm for 45s to pellet the beads. Any beads remaining trapped at the aqueous/paraffin interface were dislodged by gentle stirring with a sterile pastette before respinning the tube as before. The supernatant was removed without disturbing the pellet and the beads then resuspended in 50 ml of PBS. The beads were pelleted again as before, this time they were resuspended in a small volume of PBS before transferring them to a fresh tube and topping up the tube with PBS. They were pelleted and resuspended four more times to remove all traces of paraffin from the solution. The beads were then resuspended in 50 ml LiDS solution ( 1% lithium dodecyl sulphate, 100mM EDTA, 10 mM Tris HCl pH8.0 : the EDTA/Tris solution was autoclaved before adding filter-sterilised LDS solution). The beads were incubated at room temperature for 2 minutes to allow the LiDS solution to diffuse into the beads. The microbeads were then pelleted in Sorvall RT6000 at 3000rpm for 5 minutes and resuspended in a fresh 50ml of LiDS, before incubating at room temperature for 20 minutes to allow lysis of the agarose-encapsulated cells. The beads were pelleted and reincubated in fresh LiDS solution five more times to ensure complete lysis. The agarose microbeads were then washed in 50 mM EDTA pH8.0 several times, incubating the beads for 20 minutes each time to allow for diffusion of the solutions, until no traces of detergent remained. The beads were stored in 50mM EDTA pH8.0 at 4°C until required.

### **2.7.8 Restriction digestion of microbeads for pulsed field gel electrophoresis**

Beads which have been stored were spun in a Sovall RT6000 maximum speed for 5 min, and the supernatant carefully removed. 750µl of this slurry was withdrawn per digest to be done into a conical-bottomed Sterilin tube. This was made up to 10ml with sterile water and mixed on a tube rotator for 20 min. at room temperature. The microbeads were then pelleted in a Wifuge at 2000rpm for 5 min, the supernatant was removed and the beads resuspended in a further 10ml of water. This washing procedure was repeated twice more. The final pellet of beads was divided equally between a number of Eppendorf tubes, one for every digestion reaction to be performed. The beads were then pelleted in a microcentrifuge at top speed for one min: the volume of beads should be approximately 150-200µl per tube. For a typical restriction digestion reaction we added; 25µl of the appropriate restriction buffer at 10x concentration, 2.5µl BSA (20mg/ml), 20 units of restriction enzyme and made the reaction mix up to 400µl with water + 0.01% Triton-X100. The tube was mixed for a few seconds on a vortex to resuspend the beads in the reaction mix and the tube was then incubated at the appropriate temperature for 2hr. For digests with I-SceI, 25µl of 10x buffer, 25 units of restriction enzyme, and 4µl 20ng/µl stabilising oligo (B.M.) were mixed with the beads and then incubated on ice for 100 min to allow the oligo, the enzyme, and buffer to diffuse into the beads. The stabilising oligonucleotide contains a sequence sufficiently close to that of the recognition site of the enzyme for it to be bound by the enzyme but not close enough for it to be cleaved by it. The oligo therefore acts to stabilise the I-SceI restriction enzyme in the presence of magnesium ions and thus prolong its activity. 25µl of 50mM magnesium chloride was then added and the tube was gently mixed before incubating at 37°C for 1hr. To stop the reaction the tubes were put on ice for a few minutes, 1ml 1X TE was added, and the beads were gently mixed. They were then pelleted in a microfuge before removing the supernatant and resuspending them in a fresh ml of TE. This washing step was repeated once more. The reactions were stored at 4°C in 1.5ml TE until they were ready to load on to the gel.

## **2.8 Electrophoresis**

### **2.8.1 Agarose gel electrophoresis**

Digested genomic DNA was electrophoresed overnight on a 0.8% agarose (Sigma agarose grade II) gel in 1x TBE (0.089M Tris Borate, 0.002M EDTA, pH8.0) at 50 volts. Clearer hybridisation results were obtained from blots of gels made up with 1x TAE (0.04M Tris acetate, 0.001M EDTA, pH 8.0) which had been run at 70v for 4hr



in 1x TAE, changing the running buffer three times during the electrophoresis. Where DNA fragments in the range (0.5-6kbp) were to be resolved a 1-1.5% gel was used. For fragments under 500bp (for instance with PCR products) a 2.5% or 1.5% Nuseive (FMC Bioproducts) with 1% Sigma agarose gel was used in 1x TBE buffer, and the gel was run at 80v for ~1hr. Ethidium bromide was added to all agarose gels at a concentration of 0.5µg/ml. One tenth of the sample volume of loading buffer was added to the DNA samples prior to loading on the gel. The size markers used were Hae III-digested ΦX 174 DNA(Promega) (size range 72-1300 bp) or Hind III digested bacteriophage λ DNA (size range 0.1-23 kbp) (BRL), or 1kb DNA ladder (size range 0.1-12 kbp)(BRL). Resolved DNA fragments were visualised on a transilluminator and photographed using a video copy processor (Mitsubishi), or on Kodak film.

### **2.8.2 Preparative agarose gel electrophoresis**

DNA fragments were isolated by running them out on a low melting point agarose gel (Ultrapure LMP agarose, BRL) in 1x TAE (0.04M Tris acetate, 0.001M EDTA, pH 8.0) at 4°C. Concentrations of between 1-1.4% agarose were used depending on the size of the fragment to be resolved. The gels were viewed on a transilluminator and the specific fragment was cut out from the gel using a sterile scalpel blade. Care was taken to insure that the minimum size of gel slice was isolated. DNA was extracted from the gel slice by digestion of the agarose with agarase (B.M.) according to the manufacturers instructions, followed by phenol:chloroform, chloroform extractions and ethanol precipitation (as described in 2.3.1). The pellet was resuspended in a small volume (20µl) of 1x TE.

### **2.8.3 Pulsed Field Gel Electrophoresis**

This protocol was taken from Maule (1994).

The digested beads were loaded into 240mm<sup>2</sup> wells in a 250ml 1% agarose gel made with 0.5x TAE. The yeast marker plugs containing DNA from *Saccharomyces cerevisiae* (strains AB970, YP148) and *Schizosaccharomyces pombe* (strain 3B3) (gifts from J. Maule) were equilibrated in 0.5x TAE for 1 hour before loading onto the gel. The beads and plugs were sealed into the wells with 0.5% LMP agar. Electrophoresis was carried out in CHEF DRII apparatus (J. Maule 1993). The tank was filled with running buffer (0.5x TAE) so that it was at a level of 3mm above the gel and it was kept at a constant temperature of 10°C by circulation through a cooling apparatus.

Pulse times and running voltage depended on the size of DNA fragments that were being separated.

After electrophoresis, the DNA was visualised by staining with ethidium bromide for 20 mins and photographed.

## 2.9 Transfer of DNA to membranes

### 2.9.1 Southern transfer of DNA

DNA which had been separated on an agarose gel was transferred onto a nylon membrane (Hybond-N, Amersham) by the capillary blotting method described by Southern (1975). After photography of the gel, it was denatured for 2x 20 min in 0.5M sodium hydroxide, 1.5M sodium chloride, and then neutralised for 45mins in 2M sodium chloride, 1M Tris HCl, pH5.5. An additional step was included before denaturation when blotting pulsed field gels. To aid the transfer of the large DNA molecules separated by PFGE, gels were soaked in 0.25M hydrochloric acid for 20 min. This partially depurinates and fragments the DNA. Gels were placed on 3MM blotting paper soaked in 20x SSC which acted as a wick from the reservoir of 20x SSC below. A pre-wetted nylon filter was placed onto the gel and then was covered by two pieces of 3MM blotting paper also pre-soaked in 2x SSC. A ~6cm stack of paper towels was layered on top of this and a glass plate and a 500g weight were placed on top. Pulsed field gels were left to blot for 2 days. Standard genomic DNA gels were transferred overnight. Transfer of digested plasmid DNA was for not less than 4hr.

After blotting, the nylon filters were rinsed briefly in 2x SSC. DNA was cross-linked to the filter by exposure to UV (1200  $\mu$ Joules) in a stratalinker (Stratagene) and then baked for 30 min at 68°C.

### 2.9.2 Colony transfer

This method was used to screen large numbers of bacterial colonies for recombinant plasmids (as described in Maniatis *et al.* 1989).

An L-agar plate spread with a bacterial culture was incubated at 37°C until colonies were visible, usually overnight. The plate was then chilled at 4°C for 30-60 min to harden the colonies. A labelled, gridded nitrocellulose filter was placed on top of the plate, grid side down for 30s; if duplicate lifts were required a second filter was left on the plate for 1 min. Orientation marks were made using a sterile hypodermic needle. The filter was then carefully removed from the plate and placed, colony side up, on a sheet of Whatman 3MM paper soaked in 10% SDS for 3 min. After this, the filter was transferred to a sheet of 3MM soaked in denaturing solution (0.5M sodium hydroxide, 1.5M sodium chloride) for 5 min and then to a sheet of 3MM soaked in

neutralising solution (1.5M sodium chloride, 0.5M Tris HCl pH 7.2, 0.01M EDTA) for 5 min. The filter was then rinsed in 2 x SSC for 5 min before being allowed to air-dry for 30 min. It was then baked at 80°C for 1-2 hours under vacuum. The plate from which the lift had been taken was incubated at 37°C for several hours to allow the colonies to regrow.

## 2.10 Radiolabelling of DNA

### 2.10.1 Random Priming of DNA Probes

Probes were labelled with [ $\alpha$ -<sup>32</sup>P] dCTP by random priming from hexadeoxyribonucleotides using the Klenow fragment of *E. coli* DNA polymerase I (Feinberg and Vogelstein 1983, 1984)

Either a 20 $\mu$ l aliquot of gel isolated probe or 50ng of DNA in an 11 $\mu$ l volume was denatured by boiling for 5 min. It was then cooled rapidly on ice or if a gel derived probe, at 37°C. A random priming kit (B.M.) was used to label the probe: 2 $\mu$ l of 10x reaction buffer, 1 $\mu$ l each of dATP, dGTP and dTTP, 2 units of Klenow and 30 $\mu$ Ci of [ $\alpha$ -<sup>32</sup>P] CTP were added to the denatured probe. The reaction was incubated at 37°C for a minimum of 45 min. Unincorporated nucleotides were removed by running the reaction through a Sephadex G-50 Nick column (Pharmacia), according to the manufacturers instructions. The probe was denatured by heating to 100°C for 10 mins in the presence of 100 $\mu$ l 10mg/ml sonicated salmon sperm DNA and then snap-cooled on ice for 10 min before adding to the hybridisation mix.

### 2.10.2 End-labelling of DNA oligonucleotides

Oligonucleotides were labelled by the transfer of the <sup>32</sup>P-labelled  $\gamma$ -phosphate from [ $\gamma$ -<sup>32</sup>P] ATP onto the terminal 5'-OH group using polynucleotide kinase (Feinberg and Vogelstein 1983 & 1984). 30ng of DNA was labelled in a 20 $\mu$ l volume containing 1x bacteriophage T4 polynucleotide kinase buffer (5mM Tris pH 7.5, 1mM magnesium chloride, 0.5 mM DTT), 10 units of T4 polynucleotide kinase (B.M.) and 30 $\mu$ Ci of [ $\gamma$ -<sup>32</sup>P] ATP(Amersham). The reaction was incubated for 30-40mins at 37°C. It was added to the hybridisation mix without the need for removal of unincorporated nucleotides.

## 2.11 Hybridisation of membranes

### 2.11.1 Hybridisation solutions

*100x Denhardts solution*

2% BSA, 2% polyvinyl pyrrolidone, 2% Ficoll, 1mM EDTA. Stored at 4°C.

### *Hybond hybridisation mix*

6x SSC, 10% dextran sulphate, 0.1% sodium pyrophosphate, 0.1% sodium dodecyl sulphate (SDS), 5x Denhardt's solution.

### *Oligo Hybridisation Mix*

6x SSC, 4x Denhardt's solution, 0.1% sodium pyrophosphate, 0.1% SDS

### *Sonicated 'salmon' sperm DNA*

1g of herring sperm DNA (Sigma) was dissolved in 100ml of distilled water overnight. It was sonicated until fragment sizes of the DNA ranged between 300-700bp. It was stored at a final concentration of 10 mg/ml at -20°C.

## **2.11.2 Pre-hybridisation**

Pre-hybridisation and hybridisation of filters to be probed with end labelled oligonucleotides was carried out in oligo hybridisation mix at 5°C below the  $T_m$  of the oligonucleotide. For those to be hybridised with random primed or nick-translated probes Hybond hybridisation mix was used.

The Hybond filter to be hybridised was placed in between 2 nylon gauze membranes whilst in a solution of 2x SSC. The filter and membranes were rolled up, taking care to exclude air bubbles, and placed in a Hybaid hybridisation bottle. 100µl 10mg/ml sonicated salmon sperm DNA was boiled for 8 min and snap cooled on ice for 3 min before adding to 20 ml pre-warmed hybridisation mix. This solution was added to the membranes in the bottle which was then incubated with continuous rotation for at least 3 hr (overnight for new filters) in a Hybaid hybridisation oven.

## **2.11.3 Hybridisation**

End-labelled oligonucleotides were added directly to the bottle after prehybridisation, aliquotting the probe carefully into the oligo hybridisation mix, avoiding direct contact with the membranes. Hybridisation was carried out for at least 4hr, at the same temperature as the prehybridisation. Filters were then removed from the bottles and washed at room temperature in 4x SSC, 0.1% SDS, and then again in the same solution heated to the temperature of the hybridisation reaction.

1mg of denatured, sonicated DNA was added to the denatured random primed probe before it was added to the hybridisation mix. The hybridisation reaction was incubated overnight at 68°C. The filter was washed at 68°C in 2x SSC, 0.1% SDS, and then at 0.2x SSC, 0.1% SDS, each for 20 min. If the background level of radioactivity on the filters was still high, they were washed in 0.1x SSC, 0.1% SDS for a further 20 min.



#### **2.11.4 Detection of hybridization**

##### *Autoradiography*

Filters were exposed to Kodak X-OMAT film in cassettes with intensifying screens. Filters hybridised to  $^{32}\text{P}$ -labelled probes were exposed at  $-70^{\circ}\text{C}$ . The films were stored between 30min and two weeks before being developed using an automatic Fuji X-ray film processor RGII.

##### *Phosphor-imaging*

Filters were exposed to a storage phosphor screen for between 30min and 48hr. The screens were then scanned on a Molecular Dynamics PhosphorImager. The image was displayed within a grey scale on a computer screen using Image Quant Molecular Dynamics software, which allowed adjustments to be made to improve visualisation of the signal before it was printed out by a laser printer.

#### **2.11.5 Re-using filters**

Radioactive probes were stripped from blots which had been previously hybridised to enable them to be used again. This was done by incubating the filters in 0.4M sodium hydroxide at  $45^{\circ}\text{C}$  for 30 min, then neutralising them in 0.1x SSC, 0.1% SDS, 0.2M Tris pH 7.5 for 15 min. Alternatively, membranes were immersed in 500ml 0.1% SDS which had been heated to  $100^{\circ}\text{C}$ , for 10 min. The filters were checked by detection as described above.

### **2.12 Sequencing of DNA**

#### **2.12.1 Double strand DNA sequencing**

This protocol was taken from B. Stillman, Cold Spring Harbor (unpublished).

Template plasmid DNA was prepared according to the miniprep protocol and was finally resuspended in 50 $\mu\text{l}$  1x TE. To 9 $\mu\text{l}$  of the suspension 1 $\mu\text{l}$  of RNase A (50mg/ml) was added and the mixture was incubated for 15 min at room temperature. 1 $\mu\text{l}$  of sequencing primer (10ng/ $\mu\text{l}$ ) and 1 $\mu\text{l}$  of 1M sodium hydroxide were then added and the mixture was left to incubate at  $68^{\circ}\text{C}$  for 10 min to denature the template DNA in the presence of the primer. The annealing reaction was allowed to occur in the presence of 4 $\mu\text{l}$  of TDMN (0.3M TES (Sigma T-1375), 0.05M DTT, 80mM magnesium chloride, 0.2M sodium chloride, 1% chloroform, pH 1.6, stored at  $4^{\circ}\text{C}$ ) over a 10 min incubation period at room temperature, before a brief incubation on ice. To this reaction 1 $\mu\text{l}$  0.1 M DTT, 0.4 $\mu\text{l}$  5x Sequenase (USB) labelling mix (7.5 $\mu\text{M}$  each dGTP, dCTP, dTTP), 2.1 $\mu\text{l}$  LoTE (3mM Tris HCl pH 7.5, 0.2 mM EDTA), 0.5 $\mu\text{l}$  [ $\alpha^{35}\text{S}$ ] dATP (Amersham) and 2 $\mu\text{l}$  T7 DNA polymerase (Sequenase (USB) diluted 1:7 with enzyme dilution buffer, USB) was added. The extension reaction was

incubated for 5 min at room temperature. 4µl of the reaction was then added to each of 2.5µl of the dideoxynucleotide termination mixes (containing ddATP, ddGTP, ddTTP or ddCTP, USB) which had been pre-warmed at 37°C for 5 min. The termination reactions were incubated for 10 min at 37°C before the addition of 5µl STOP mix (USB). The reactions were stored at -20°C before subjecting them to polyacrylamide gel electrophoresis.

### **2.12.2 Polyacrylamide Gel Electrophoresis**

Denaturing polyacrylamide gels were used to visualise DNA sequencing reactions. The glass plates were prepared by washing them thoroughly with detergent and rinsing with water and then 100% ethanol. The back plate was coated in dimethyldichlorosilane solution, and the front plate in 10ml absolute ethanol containing 30µl glacial acetic acid and 30µl  $\gamma$ -methacryloxypropyltrimethoxysilane. This was to ensure that when the plates were separated after electrophoresis that the gel would adhere only to the back plate. The plates were placed together with the treated sides next to each other separated by spacer strips before being taped together. 30 ml 6 % acrylamide (40% deionised acrylamide, 4M urea, 1x TBE stored in the dark at 4°C) was combined with 30µl N,N,N',N'-tetramethylethylenediamine (TEMED) and 180µl ammonium persulphate (10% in water freshly made up). This was immediately poured into the gel cast. A well-former was inserted into the top of the gel before clamps were applied whilst the gel set.

Samples were denatured at 80°C for 2 min before they were loaded onto the gel. Denaturing gels were run in 1x TBE on BRL vertical apparatus, model II at 26 watts (~1500v) for 1-2 hours. After electrophoresis, the plates were separated leaving the gel attached to the back plate. The gel was fixed in 10% methanol, 10% glacial acetic acid for 20 min and then rinsed in water for 15 min before being dried at 80°C for 30 min.

## **2.13 Polymerase Chain Reaction**

### **2.13.1 Oligonucleotide preparation**

Oligonucleotides were synthesised, in the form of ammonium stocks, on an Applied Biosystems 381A oligonucleotide synthesizer, by A. Gallhager. Oligonucleotides were precipitated from stocks by the addition of 1/10 volume 3M sodium acetate and 2.5 volumes ethanol, and incubating for 1 hour at -20°C. They were pelleted by a 15 minute centrifugation at 13.000g, washed twice in 80% ethanol and dried under vacuum. The pellet was resuspended in 200µl TE and the concentration determined



by measurement of the optical density where  $OD_{260} = 1$  at 25µg/ml single stranded DNA

**2.13.2 PCR with Genomic DNA**

Reactions were carried out in a 50µl volume which contained 5µl 10X buffer (500mM potassium chloride, 100mM Tris HCl pH 9.0, 1% Triton X-100 (Promega)), 3µl 25mM magnesium chloride (Promega), 1µl of each primer at 250ng/ul, 1µl of 50X dNTP stock to give a final concentration of each dNTP of 200µM, approximately 100ng of template DNA and 1 unit of thermostable DNA polymerase (Promega). For non-preparative amplification, reactions were carried out in 25µl volumes. Reactions were overlaid with mineral oil (Sigma). All PCR reactions were carried out on a Hybaid omnigene PCR apparatus.

PCR conditions were determined based on the G:C content of the oligonucleotides, the length of the product expected, the specificity of the oligonucleotides, and were optimised empirically. Each set of conditions always had an initial cycle with a longer denaturation step at a higher temperature which was followed by a series of cycles, the number of which was dependent on the efficiency of amplification (generally 30 cycles were performed). These cycles consisted of a denaturation step (usually 91°C for 30s followed by an annealing step. The annealing temperature was usually 5°C lower than the melting temperature ( $T_m$ ) of the primers, which is determined by the equation  $T_m (1M [Na^+]) = 4(G+C) + 2(A+T)$ . Annealing was carried out for 30 seconds. It was followed by an extension step of 1 min per kb of DNA to be amplified at 72°C. The final cycle had an increased extension step of 5 or 10 minutes depending upon the size of the amplified fragment. To improve the specificity of the reaction conditions for 'touch-down' PCR were used, as described in Don *et al.* 1991.

10µl of the product was visualised on an ethidium bromide-stained agarose gel.

**2.13.3 Conditions and primers for PCR analysis of targeted cell lines**

*Neomycin*

Oligonucleotide 673    5' GCG ATG CCT GCT TGC CGA 3'    18 Mer

Oligonucleotide 670    5' GAA GGC GAT AGA AGG CGA 3'    18 Mer

Reaction conditions:

- 93°C: 1 min, x 1
- 60°C: 1 min, 72°C: 30s, 91°C 30s: x 10
- 60°C: 1 min, 72°C: 1 min, 91°C 30s: x 10
- 60°C: 1 min, 72°C: 1.5 min, 91°C 30s: x 10

60°C: 1 min, 72°C 5min: x 1

Expected product size: ~212bp

### *Leu2*

Oligonucleotide 397 5' GAA GTC GGT GAT GCT GTC 3' 18 mer

Oligonucleotide 399 5' TGT GTG GTG CCC TCC TCC 3' 18 mer

Reaction conditions ('Touchdown' PCR):

93°C: 2 min; 70°C: 45s; 72°C: 45s: x1

91°C: 30s; 68°C: 45s; 72°C: 45s: x2

91°C: 30s; 64°C: 45s; 72°C: 45s: x2

91°C: 30s; 62°C: 45s; 72°C: 45s: x2

91°C: 30s; 60°C: 45s; 72°C: 45s: x2

91°C: 30s; 58°C: 45s; 72°C: 45s: x20

91°C: 30s; 58°C: 45s; 72°C: 5 min: x1

Expected product size: ~300bp

### *WT1*

Oligonucleotide C911 5' ACT TCA CTC GGG CCT TGA TAG 3'

Oligonucleotide C912 5' GTG GAG AGT CAG ACT TGA AAG 3'

Reaction conditions:

94°C: 2 min, 57°C: 15s, 72°C: 1 min: x 1

92°C: 30s, 57°C: 15s, 72°C: 1 min: x 30

92°C: 30s, 57°C: 15s, 72°C: 10 min: x 1

Expected product size: 275bp

### *DIIS533*

Oligonucleotide C912 5' CTC TGC CTA GTC CCT GGG TG 3' 20 mer

Oligonucleotide C911 5' TGG GGG TCT GGG AAC ATG 3' 18 mer

Reaction conditions ('Touchdown' PCR)

94°C: 2 min, 65°C: 15s, 72°C: 30s: x 1

91°C: 30s, 68°C: 15s, 72°C: 30s: x 2

91°C: 30s, 66°C: 15s, 72°C: 30s: x 2

91°C: 30s, 64°C: 15s, 72°C: 30s: x 2

91°C: 30s, 62°C: 15s, 72°C: 30s: x 2

91°C: 30s, 62°C: 15s, 72°C: 5 min: x 1

Expected product size: ~450bp

## 2.14 Fluorescence in situ hybridisation (FISH)

The protocols for fluorescence in situ hybridisation have been taken and modified from Gosden (pers. comm.), Trask (1991), Lichter and Cremer (1992), Y. Shibaski (Ph.D thesis 1994).

### 2.14.1 Metaphase Spreads from cell cultures

A confluent culture of cells was harvested from a 25ml flask and was used to reseed a fresh 25ml flask in a dilution of 1 in 5 in a total of 5ml of fresh medium. The flask was incubated overnight, or until the cells reached approximately 60% confluence. Fifty microlitres of colcemid (0.01mg/ml) was then added, and the flask incubated for a further 45 min. The cells were trypsinised as described in section 2.6.2, then the spent medium and PBS rinse were harvested along with the trypsinised cells into a 10ml conical-bottom Sarstedt tube. The cells were pelleted by centrifugation at 1200rpm for 5 min and the supernatant was then removed. The pellet was resuspended in 5ml hypotonic solution (0.28% potassium chloride, 0.5% trisodium citrate), and incubated at room temperature for 15 min. The cells were centrifuged as before and the supernatant was removed. The pellet was resuspended in a fresh fixative solution of 3:1 methanol: glacial acetic acid, adding the first ml of solution dropwise whilst vortexing. The tube was stored for at least 24 hr or up to 5 days at 4°C.

To make metaphase spreads on glass slides the cells had to be washed twice in 5ml of fresh fixative. They were finally resuspended in a small volume of fixative; so that the cell suspension appeared just cloudy. A drop of the suspension was allowed to fall from arm's length onto acid-cleaned glass slides which had just been moistened with condensed breath. When the suspension had air-dried the slides were checked under a microscope to locate metaphases free of cytoplasm which were required for clean *in situ* hybridisation. Tubes containing the suspensions of metaphase were refilled with fresh fixative and stored at -20°C where they remained usable for up to two months.

### 2.14.2 Preparation of extended chromosomes and a high density of metaphases

Cultures were grown to 60% confluence as for preparation of regular metaphase chromosomes. One hundred microlitres of 0.01mM methotrexate ('block') was then added to the medium on the cells and the flask was incubated for a further 18 hr. Methotrexate is a folate analogue and acts to inhibit DNA synthesis so most of the cells are arrested in mid-S phase. After 18 hr the cell monolayer was rinsed twice with PBS to remove the methotrexate and then 5 ml of fresh medium was added. This should stimulate a synchronous entry into mitosis by most of the arrested cells. The cells were incubated for between 4 to 6 hr before 50 µl of colcemid was added,

then incubation was continued for a further 30 min. After this time the flask was rapped vigorously sideways a couple of times to dislodge dividing cells, which were then harvested in the spent medium. The monolayer was rinsed with PBS which was also harvested. The rest of the monolayer was discarded as it contains mainly non-mitotic cells. The cell suspension was then treated exactly as for regular metaphases to produce slides for *in situ* hybridisation.

### **2.14.3 Production of digoxigenin- or biotin-labelled probe by nick-translation**

The probes used for FISH were; pTVLO $\Omega$ , a derivative of the plasmid pTVLO which does not contain human repeat sequences (see 3.1.1); cosINSUL/IGF2 ('IGF2'), a cosmid containing parts of the genes for tyrosine hydroxylase, insulin, and insulin growth factor receptor 2, all located on chromosome 11p15.5 (de Pagter-Holthuisen *et al.* 1987); pHS53, a plasmid containing a cloned alphoid repeat element specific for human chromosome 11 (Yurov *et al.* 1987), pYAM7-29, a plasmid containing an alphoid repeat specific for human chromosome 17 (Yurov *et al.* 1987).

The haptens used to label the probes were either biotin bound to dUTP (Bio-dUTP) or digoxigenin bound to dUTP (DIG-dUTP). DIG-dUTP is supplied by B.M. at a concentration of 500 $\mu$ M in a mix with dTTP where DIG-11-dUTP : dTTP = 3:1, and Bio-dUTP is supplied at a concentration of 500 $\mu$ M by B.M. as Bio-16-dUTP, or by Sigma as Bio-11-dUTP.

Two micrograms of plasmid DNA in 10 $\mu$ l of water was combined with 10 $\mu$ l 10x nick buffer (500mM Tris HCl pH 8.0, 50mM magnesium chloride, 500 $\mu$ g/ml BSA, 100mM  $\beta$ -mercaptoethanol), and 10 $\mu$ l of 10x dNTP mix (500 $\mu$ M each dATP, dGTP, dCTP). Then either 10 $\mu$ l 500 $\mu$ M DIG-11-dUTP, or 10 $\mu$ l 500 $\mu$ M Bio-11-dUTP was added, with 20 units of DNA polymerase I (BRL) and 1 $\mu$ l of a 1/500 dilution of DNase I (B.M.) and the reaction was brought up to a total volume of 100 $\mu$ l with water. The reaction was incubated at 15°C for 2-3 hr. The probe length was roughly estimated at this stage (to check that DNase I activity was sufficient) by boiling a 6 $\mu$ l aliquot of the reaction for 3 min, snap cooling it on ice then loading it directly on to a 2% agarose gel. The size range of the fragments was visualised by staining with ethidium bromide. If the sizes ranged above 800bp more DNaseI was added and the incubation time was extended. Five microlitres of 500mM EDTA was added to stop the reaction.

Unincorporated nucleotides were removed by passing the reaction mix through a 1ml Sephadex G-50 spin column pre-equilibrated with TES (10mM Tris HCl (pH 8.0), 1mM EDTA, 20mM sodium chloride, 0.1% SDS). The column was centrifuged at 1500g for 2 min to collect 100 $\mu$ l purified probe. If the volume had increased, the

DNA was precipitated with 2 vol. ethanol, 1/10 vol. 3M sodium acetate, 10 $\mu$ l 10mg/ml sonicated salmon sperm, and resuspended in 100 $\mu$ l 1x TE.

#### **2.14.4 Check of probe incorporation of label**

Serial tenfold dilutions down to 10<sup>-5</sup> of purified probe were spotted onto a nitrocellulose filter. A labelled standard concentration of DNA was also spotted onto the filter, when available. The filter was UV cross linked and then baked at 68°C for 20 min. After baking, the filter was soaked for 5 min in buffer A (0.1M Tris HCl pH 7.5, 1x SSC), then incubated in buffer B (0.1M Tris HCl pH 7.5, 1x SSC, 3% BSA) for 30 min at room temperature. Five microlitres (4U) of anti-DIG or anti-biotin antibodies conjugated to alkaline phosphatase (B.M.) in 5ml buffer B was made. The filter was incubated in this solution in a sealed plastic bag for 30 min at room temperature. The filter was then washed for 3 x 5 min in buffer A and then 1x 5 min in buffer C (0.1M Tris HCl pH 9.5, 1x SSC, 50 mM magnesium chloride). Two drops of each of the solutions in an alkaline phosphatase-detection kit (Vector Labs) was added to 5ml buffer C. The filter was incubated in the dark in this solution sealed in a plastic bag for several hours until the colour developed. The filter was rinsed in buffer A and allowed to air dry before storage.

#### **2.14.5 Check of probe length**

The size of the labelled probe fragments was verified by running a 5 $\mu$ l aliquot in 20 $\mu$ l alkaline loading buffer (ALB) (50mM sodium hydroxide, 1mM EDTA, 2.5% Ficoll, 0.025% bromocresol green) on a 14cm alkaline gel. The alkaline gel was prepared by pouring a 1% agarose gel (50mM sodium chloride, 1mM EDTA) and allowing it to equilibrate in alkaline electrophoresis buffer (30mM sodium hydroxide, 1mM EDTA) for at least 30 min before loading. Ten to fifteen microlitres of biotin or DIG-labelled HindIII-cut  $\lambda$  DNA (BRL) in 5-10 $\mu$ l ALB was loaded alongside the probe sample. The gel was run for about 2 hours at ~40v, or until the dye front had moved about 5cm from the origin. The DNA was transferred after neutralisation to a nitrocellulose membrane by Southern blotting overnight (see 2.9.1). The DIG or biotin signal was detected as above (2.14.4). The optimal size of probe fragment is between 100-500bp.

#### **2.14.6 Direct labelling of probes with fluorochromes**

This method was essentially the same as the nick translation method used to label probes with biotin and DIG (see 2.14.3) except that the hapten-tagged nucleotide was replaced with a nucleotide bound to a fluorochrome. Ten microlitres of 1mM



rhodamine-4-dUTP or fluorescein-11-dUTP (both from Amersham) was substituted for 10 $\mu$ l 500 $\mu$ M Bio-dUTP or DIG-dUTP in the protocol described in 2.14.3. The size of the labelled fragments were estimated on a gel in the same way and unincorporated nucleotides were removed by passing the reaction through a spin column as described in 2.14.3, giving a final concentration of probe of 2 $\mu$ g in 100 $\mu$ l.

#### **2.14.7 Hybridisation with pTVLO $\Omega$ , and pH53**

Four microlitres per slide of digoxigenin-labelled pTVLO $\Omega$  (~40ng), 2 $\mu$ l/slide sonicated salmon sperm DNA (10mg/ml) sonicated to around 500bp and 2 $\mu$ l/slide bio-labelled or fluorochromo-labelled pH53 were mixed together, and precipitated with 2.5 volumes of ice cold absolute ethanol. The pellet was dried down and resuspended in 10 $\mu$ l hybridisation mix/slide (50% deionised formamide, 10% dextran sulphate, 2X SSC). Enough probe-hybridisation mix was made up for one extra slide. The mix was allowed to resuspend for at least 1 hr at room temperature with occasional gentle vortexing.

The slides had been made 3-4 days in advance and stored under light vacuum. They were then incubated in RNase (100 $\mu$ g/ml RNase A, 2X SSC) at 37°C, for 1 hr, and then dehydrated by dipping them in each of 70, 90, and 100% ethanol for 3 min each. They were then allowed to air dry.

Ten microlitres of probe/hybridisation mix was applied to an 18x18mm cover slip. The slide was inverted and pressed onto the cover slip. The coverslip was sealed onto the slide with rubber solution (Tip-Top). When the rubber solution was completely dry the slide was transferred to a thin metal tray floating in a water bath at 75°C. It was incubated for 10 min to denature the DNA. The slide was then incubated at 37°C overnight.

When using a probe labelled with a fluorochrome, care was taken to minimise exposure of the reaction to light, which would otherwise reduce the signal from the probe.

#### **2.14.8 Detection with fluorescein isothiocyanate (FITC)**

Non-specifically-bound probe was removed by washing the slides for 4 x 2 min in each of the following solutions ; 50% deionised formamide, 2 x SSC at 45°C; 2X SSC at 45°C; 0.2x SSC at 60° C. The slides were then incubated for 5 min in PN buffer (0.1M phosphate pH 8.0, 0.5% Nonidet P-40, 0.02% sodium azide) at 45°C. Forty microlitres of blocking buffer (PN, 5% non fat dry milk, 0.01% Na azide) was then applied to each slide under a 20x40mm coverslip, and the slide incubated for 5 min.



For simultaneous detection of probes labelled with DIG and biotin, anti-digoxigenin-FITC (fluorescein isothiocyanate) raised in sheep (Sigma) at 15µg/ml and avidin-FITC (Vector Labs) at 5µg/ml in 40µl blocking buffer was applied to the slides under a large coverslip, and incubated for 30 min at 37°C. (If DIG was the only hapten to be detected, avidin-FITC and biotinylated anti-avidin antibodies were omitted from amplification steps). The slides were then washed for 4 x 2min in 4x SSCT (4x SSC, 0.05% Triton X-100) at 45°C, and 5 min in PN buffer 45°C. The second antibody layer was then applied; this was anti-sheep-FITC raised in rabbit (Vector Labs) at 5µg/ml and biotinylated anti-avidin antibodies (Vector Labs) at 5µg/ml in 40µl blocking buffer per slide. The slides were incubated and then washed as before. Anti-rabbit-FITC (Vector Labs) at 5µg/ml and avidin-FITC at 5µg/ml in 40µl blocking buffer per slide was then applied. The slides were again incubated then washed as before. Finally the slides were mounted in 40 µl DAPI (4'6-diamidino-2-phenylindole)/ PI (propidium iodide) antifade solution: (50% Citifluor antifade, 50% glycerol, 0.25µg/ml DAPI, 0.5µg/ml PI), this produces an R-(reverse) banding-like staining of the chromosomes. The coverslip was sealed with rubber solution. Slides were stored in the dark at 4°C until they were observed under the confocal microscope.

## 2.15 Microscopy

### 2.15.1. Confocal microscopy

Samples were viewed initially under Ultraviolet light using a Biorad confocal microscope to identify metaphases which showed specific hybridisation signal. These metaphases were then scanned using a laser directed through the confocal microscope. The confocal laser scanning microscope contains two photomultiplier tubes which enable signals of two different wavelengths to be collected simultaneously, thus the two different wavelengths emitted by FITC and PI, when excited by the laser, can be collected at the same time. The two channels can be adjusted independently to enhance the required signal and the image is output onto a computer screen using Biorad software, improved and updated by Paul Perry (MRC HGU). The software enables the two signals to be merged so that the FITC signal from the probe is superimposed upon the PI signal from the stained chromosomes, giving a complete image of the metaphase. The merged images were stored on an optical disc.

### **2.15.2 Multicolour FISH**

A cooled CCD (charge coupled device) camera has been used to record FISH images with three different colours. This has enabled the fluorescence from FITC, rhodamine red and DAPI/PI to be visualised simultaneously.

## **CHAPTER 3**

### 3 Vector Construction

#### 3.0 Preface

All reactions were performed as described in Materials and Methods unless otherwise stated. Between each enzymatic manipulation of the constructs, proteins and salts were removed from the plasmid DNA by extracting the solution with phenol/chloroform and chloroform, precipitating with 100% ethanol and rinsing the pellet in 70% ethanol before resuspending the DNA in deionised, sterile water. To check that a digestion of plasmid DNA by restriction enzymes was complete, the DNA was run out on a 1-2% agarose gel next to an uncut sample of the same DNA and examined under UV light after staining with ethidium bromide. Ligations were made with ratios of vector DNA to insert DNA as close as possible to those described in the methods, unless otherwise stated. Transformation of bacterial cells was made by either heat shock or by electroporation of vector constructs. The highest efficiencies of transformation have been obtained by careful preparation of DH10B cells for electrotransformation. The rapid mini prep protocol was used to prepare DNA from colonies arising after transformation for the initial identification of plasmids containing the correct ligation event by restriction digest and sequencing. The mini/maxi prep method was used to produce better quality plasmid DNA for further characterisation or subsequent manipulation.

Oligodeoxynucleotides (oligos) used in the construction of the vectors and as primers in sequencing short regions of the plasmid constructs are shown in Table 3.1.

#### 3.1 Construction of pBLOB3XN<sup>-</sup>, pBLOB4XN<sup>-</sup> and pTVL1.3

See figure 3.1.0

##### 1) *pTBPNSLO*

pTBPNS1 is a conventional PNS vector with neoR and hsv-tk genes for positive and negative selection, and contains the *Leu2* gene for selection in yeast (described in detail in Brookes *et al.* 1993).

pTBPNSLO had been previously derived from pTBPNS1 by ligating pTBPNS1, linearised by a KpnI + BamHI partial digest, to annealed complementary oligonucleotides 759 and 761, which encode SmaI, SfiI, RsrII and 2x KpnI sites plus a modified lac operator sequence (Grimes *et al.* 1990). This formed the starting point for my vector construction strategy.

##### 2) *pTVLO*

pTBPNSLO was linearised by digestion with SfiI. Approximately 30ng was then ligated to ~4ng of the complementary oligos B403 and B404, which had previously

been annealed, containing HindIII, XhoI and BamHI restriction sites. After ligation, the reaction was transformed by heat shock into competent DH5α cells.

**Table 3.1 Oligonucleotides used in vector construction and sequencing**

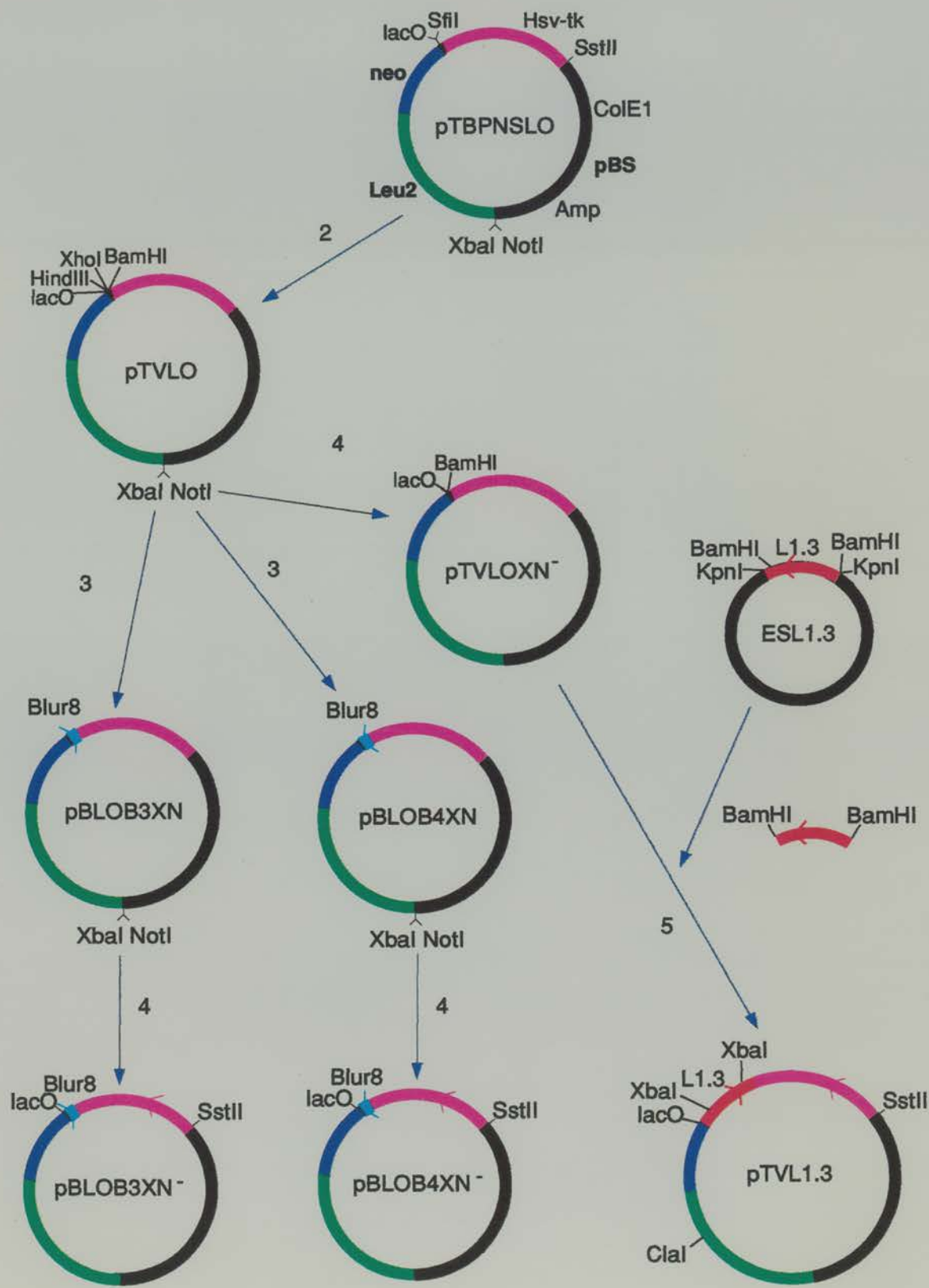
Description	Number	Oligonucleotide sequence 5' → 3'
lacO, SmaI, SfiI, RsrII	759	GATCGAATTGTGAGCGCTCACAAATT → CGGTACCGGCCGGACCGGCC
	761	GATCGGCCGGTCCGGCCGGTACCGAA → TTGTGAGCGCTCACAAATTC
HindIII, XhoI, BamHI	B403	AAGCTTCTCGAGGATCCGAC
	B404	GGATCCTCGAGAAGCTTGTC
3' end Hsv-tk	B52	GTTTTGTATTGGTCACCACG
XbaI-NotI fill-in	B787	GGCCGATCGCGA
	B790	CTAGTCGCGATC
3' end <i>Leu2</i>	B400	CAAATAAAAAACACTCAA
3' end neoR	B112	ACGGGTGTTGGGTCGTTTGTTCGG
Universal primers	291	AACAGCTATGACCATG
	292	GTAAAACGACGGCCAGT
I-SceI	B72	GATCCTAGGGATAACAGGGTAAT
	B73	GATCATTACCCTGTTATCCCTAG

Twelve colonies were picked from those produced by the transformation, and DNA was isolated from them by the rapid miniprep method. The DNA was digested with BamHI and run out on a gel to ascertain which plasmids were the products of the required ligation event and which therefore contained a unique BamHI site supplied by the oligos. One clone identified in this way was selected and a short section of it was sequenced by the rapid protocol, using the oligo B52 as a primer. The presence of all required sites was verified. This clone was named pTVLO.

**3) *pBLOB3* and *pBLOB4***

pTVLO DNA was digested with BamHI. After dephosphorylation, 500ng of the linearised vector was ligated to 30ng Blur8 insert DNA. The Blur8 insert is a 280bp fragment released by BamHI digestion of pBL8. pBL8 is pBS containing two copies of the Blur8 sequence (Deininger *et al.* 1981) in tandem array (a gift from D.J. Porteous). DNA from 12 of the colonies produced by the ligation with pTVLO was

**Figure 3.1.0** Construction of pBLOB3XN-, pBLOB4XN- and pTVL1.3





digested with BamHI to determine which clone contained a 280bp insert. Four of these clones were sequenced using the oligo B52 as a primer to check the ligation event and to determine the orientation of the insert. Two clones were identified which carried one copy of the Blur8 insert in opposite orientation to each other. These were called pBLOB3 and pBLOB4.

#### **4) *pBLOB3XN<sup>-</sup>*, *pBLOB4XN<sup>-</sup>* and *pTVLOXN<sup>-</sup>***

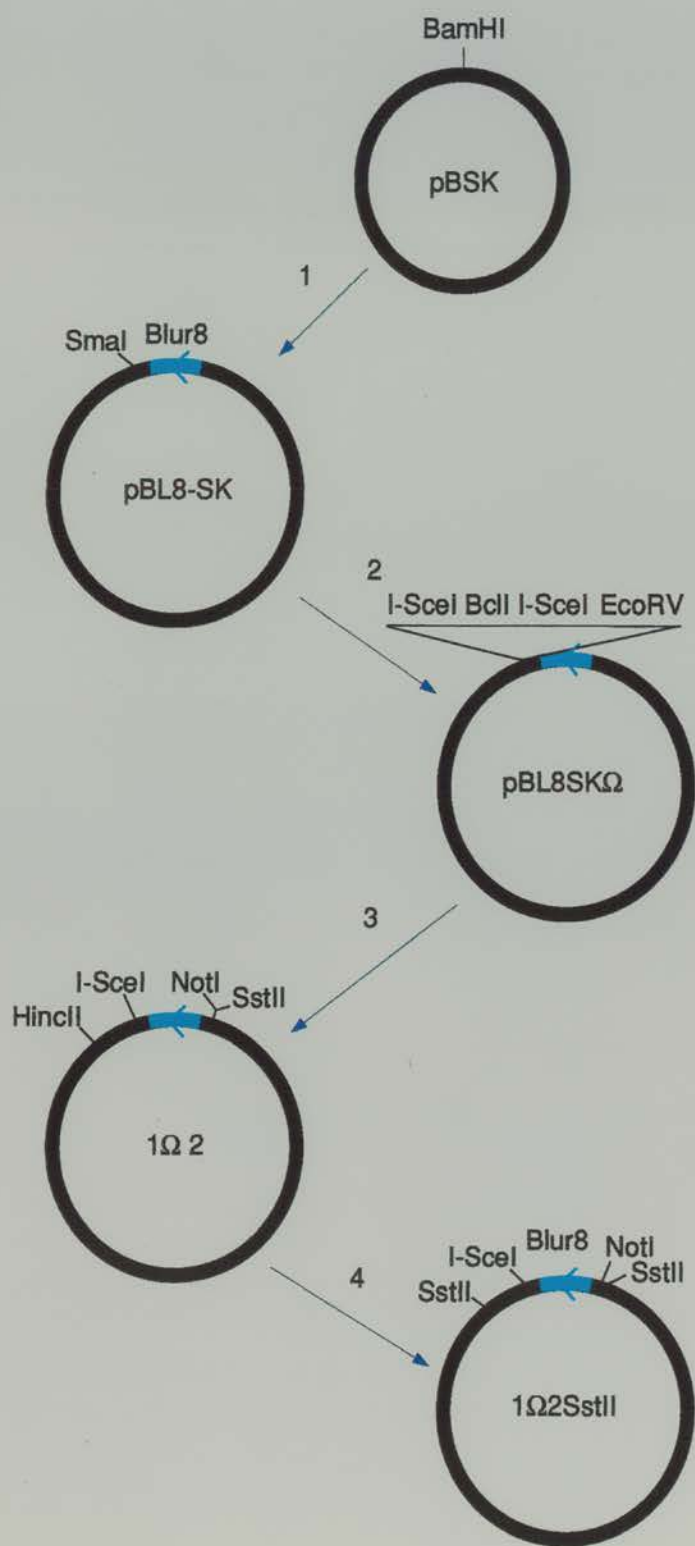
The NotI and XbaI sites in the vectors pBLOB3, pBLOB4 and pTVLO had to be destroyed. This was achieved by sequential digestion of DNA from each vector with NotI and then XbaI. 300ng of each treated vector was then ligated to the complementary annealed oligos B787 and B790. These oligos form a linker which has the appropriate sticky ends but will not recreate the restriction sites. After the ligation reaction, the DNA was again digested with NotI prior to transformation. This ensured that plasmids still containing a NotI site would not be in a covalently closed circular (ccc) form and would therefore only transform with negligible efficiency, thus enriching for transformants in which the NotI site had been destroyed. DNA prepared by mini prep from colonies arising from the transformation was subjected to digestion with XbaI to identify successful ligation events with each of the vectors. A short region from one of each type of clone was then sequenced using the primer B400 to ensure that XbaI and NotI sites were absent. Thus pBLOB3XN<sup>-</sup>, pBLOB4XN<sup>-</sup> and pTVLOXN<sup>-</sup> were created.

#### **5) *pTVL1.3***

pESL1.3 is a pBS clone containing a KpnI fragment from the farthest 3' end of an L1 repeat identified from a KpnI genomic DNA library by E. Slorach using an oligo specific to the 3' end of L1, characterised by A.J. Brookes (unpublished). A ~1.3kb BamHI insert identified as containing most of the KpnI L1 fragment was isolated from pESL1.3. Approximately 60ng of the insert was ligated to ~180ng of pTVLOXN<sup>-</sup> DNA previously linearised by digestion with BamHI and dephosphorylated. DNA prepared from colonies arising after transformation of the ligation into DH10B cells was digested with XbaI and ClaI in single and double digests to identify the required clones. A short stretch of DNA from two clones identified in this way was sequenced using B112 as a primer. One of these had the correct insert in the required orientation and was called pTVL1.3.

Figure 3.2.0

Construction of 1 $\Omega$ 2 SstII



### 3.2 Construction of 1 $\Omega$ 2SstII

See figure 3.2.0

#### 1) *pBL8-SK*

pBSK was linearised with BamHI and dephosphorylated. Approximately 75ng of the vector was then ligated to ~7ng Blur8 insert DNA (prepared from pBL8 as described previously). The ligation was transformed into electro-competent JM83 cells (a gift from E. Emslie) and plated on HAIX plates. DNA from ten white colonies was digested with BamHI to identify those clones containing a 280bp insert. One clone was selected and digested with XbaI and PstI (enzymes which recognise sites in the polylinker of pBSK) to ensure that there was only a single insert. DNA from this clone, pBL8-SK, was sequenced over a short distance using the primer 292, to determine the orientation of the Blur8 insert.

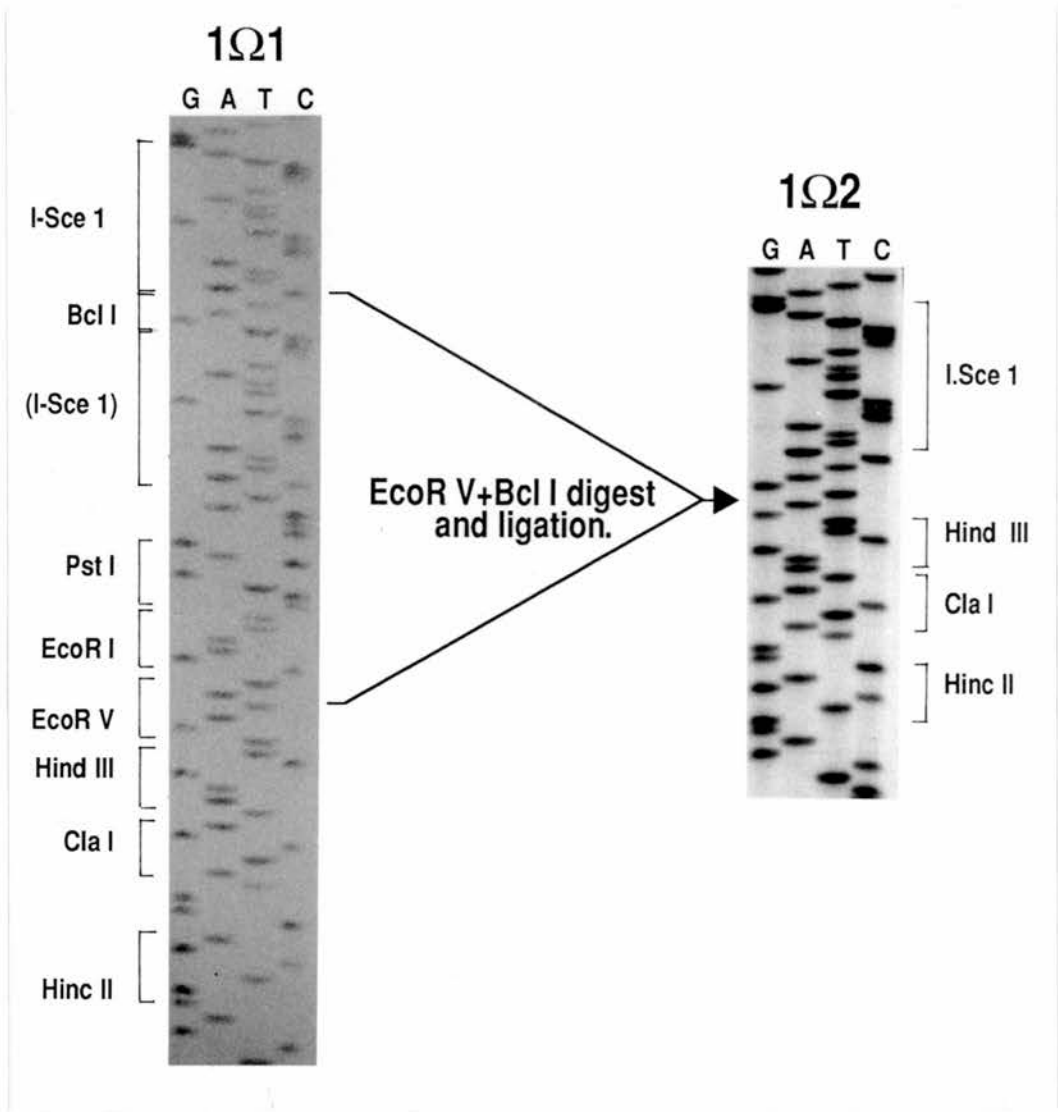
#### 2) *p1 $\Omega$ 1*

The complementary oligos B72 and B73 were annealed to produce a double stranded molecule with 5' overhangs. The double-stranded oligonucleotide was treated with T4 DNA polymerase in a 'fill in' reaction to produce a double-stranded oligonucleotide with blunt ends containing the I-SceI recognition site. pBL8-SK DNA was linearised with SmaI and ligated to the prepared oligo with a ratio of ends of vector: insert of 2:1 which required a DNA concentration ratio of 180:1. Reactions were transformed by electroporation into JM83 cells. Colony lifts were made on to nitrocellulose (see colony transfer) and the filters were hybridised to the oligo B72 (see conditions for oligo hybridisation). The filters were then exposed to X-ray film for ~ 3 hours. DNA was prepared from positively-hybridising colonies located on the original plates. The clone p1 $\Omega$ 1 was one whose DNA was linearised by digestion with I-SceI. A short region of p1 $\Omega$ 1 DNA was sequenced using the oligo 292 as a primer, and it was shown to contain two copies of the oligos B72 and B73 in tandem (see Fig. 3.2.1). Although the oligos lacked a 5' phosphate group after their initial synthesis, during the process of blunt-end filling and ligation to the vector, joining between the annealed complementary oligo ends had occurred. This had formed a structure which was stable enough to be transformed into bacteria, where the missing 5' phosphate group could be added.

#### 3) *p1 $\Omega$ 2*

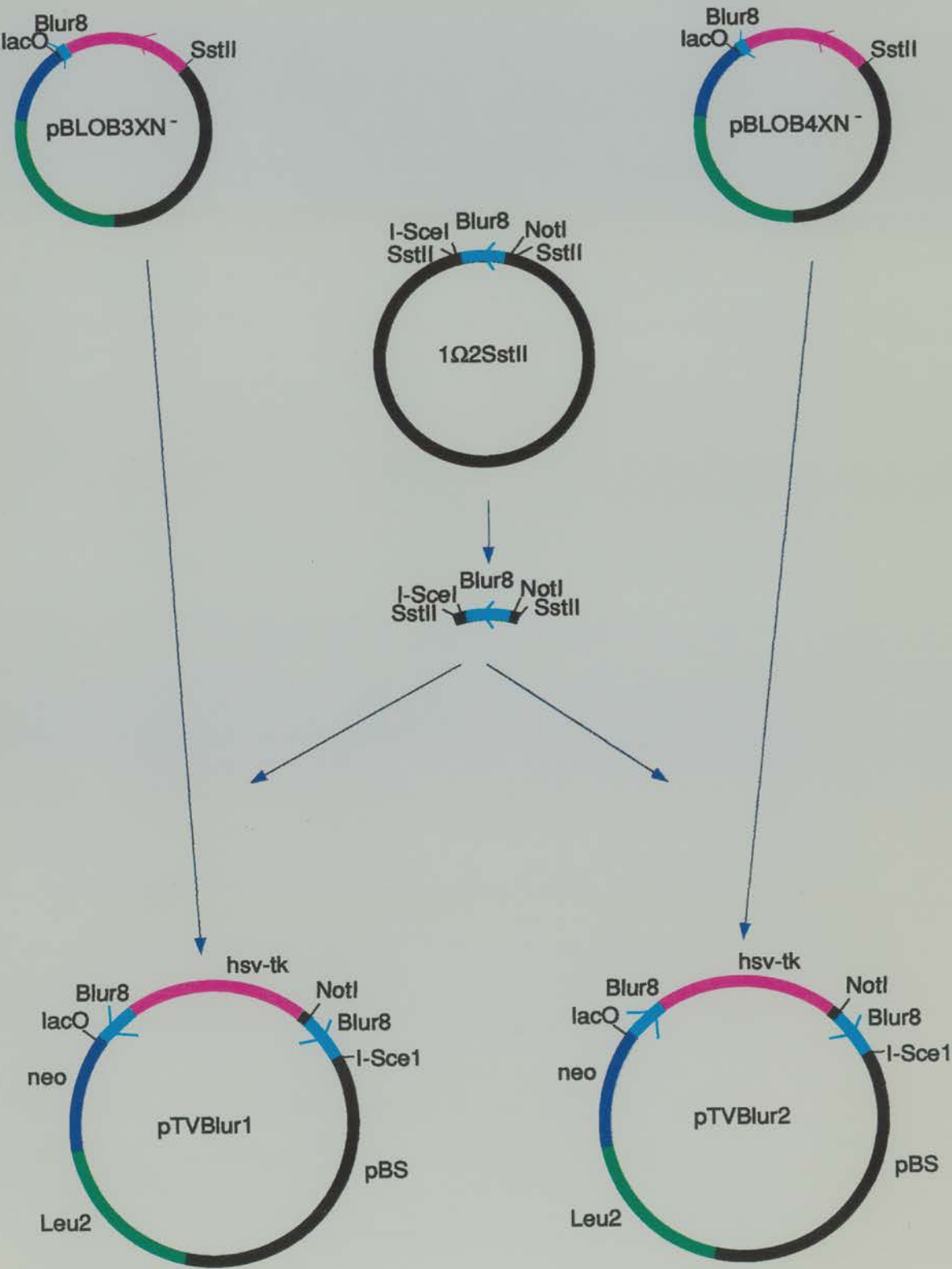
A recognition site for BclII had fortuitously been created between the two I-SceI sites in 1 $\Omega$ 1 which enabled double digestion with EcoRV and BclII followed by a blunt end ligation reaction to give the clone p1 $\Omega$ 2 containing a single I-SceI site. This was confirmed by sequencing (see Fig. 3.2.1).

**Figure 3.2.1** Sequence analysis of part of the constructs p1Ω1 and p1Ω2



DNA from the plasmids p1Ω1 and p1Ω2 was prepared by the rapid miniprep protocol described in section 2.3.1. Dideoxy-sequencing was carried out on the prepared templates exactly as described in section 2.12.1, using the universal primer 291 (see Table 3.1). The reactions were then subjected to polyacrylamide gel electrophoresis as described in section 2.12.2, for 2hr. The gel was fixed and dried before exposure to X-ray film at room temperature overnight.

**Figure 3.3.0**      Construction of pTVBlur1 and pTVBlur2





#### **4) p1Ω2SstII**

Hinc II digestion of p1Ω2 linearised the molecule, which was then dephosphorylated. 100ng of this was ligated to ~1ng blunt end phosphorylated linkers containing an SstII site (Biolabs). (SstII is an isoschizomer of SacII and KspI). The ligation was transformed into electro-competent DH5α cells. Twelve of the resulting colonies were picked and the DNA prepared from them was digested with KspI. A clone named p1Ω2SstII contained a unique SstII (KspI) fragment of ~350bp which was the insert required for the next step in the vector construction.

### **3.3 Construction of pTVBlur1 and pTVBlur2**

See figure 3.3.0

DNA from pBLOB3XN<sup>-</sup> and pBLOB4XN<sup>-</sup> was linearised with KspI and then dephosphorylated. Approximately 20 ng of the prepared vector was ligated to ~2ng of the 350bp insert fragment prepared from p1Ω2SstII (vector: insert size = 20:1). The ligation was transformed into competent JM83 cells. DNA from colonies produced was digested with HindIII to determine which clones contained the correct insert in the required orientation. Clones thus identified from both ligations were further characterised by multiple restriction enzyme digests including I-SceI, NotI and ClaI (see Fig. 3.5.1). These clones were named pTVBlur1 and pTVBlur2. A restriction map of the completed vectors is shown in Fig. 3.3.1

### **3.4 Construction of pLSst3**

See figure 3.4.0

#### **1) pBΩ2**

pBS DNA was digested with BamHI and ~800ng of this DNA was ligated to ~8ng of annealed complementary oligos B72 and B73 (see previously). After transformation into DH5α cells, the reaction was incubated on HAIX plates. White colonies produced were picked and subjected to digestion with I-SceI. DNA from two clones containing an I-SceI site were sequenced over a short distance with the primer 292, and pBΩ2 was identified as the required clone.

#### **2) pΩ1.2 8**

DNA from pBΩ2 was digested with Asp 718 (≡ KpnI) and then dephosphorylated. A 1.2kb KpnI insert was obtained from pTB1.2(11) which contains the 1.2 kb KpnI fragment from the 3' end of the β-globin locus (a gift from A.J.Brookes, modified from pBK1.2(11) described in Shafit-Zagardo *et al.* 1982). The purified insert DNA was ligated to the prepared pBΩ2 vector DNA in a reaction containing 100ng of each type of molecule (vector: insert size = 2:1). After electro-transformation of DH5α cells,

**Figure 3.3.1**      Restriction maps of the vectors pTVBlur1 and pTVBlur2 (approximately to scale)

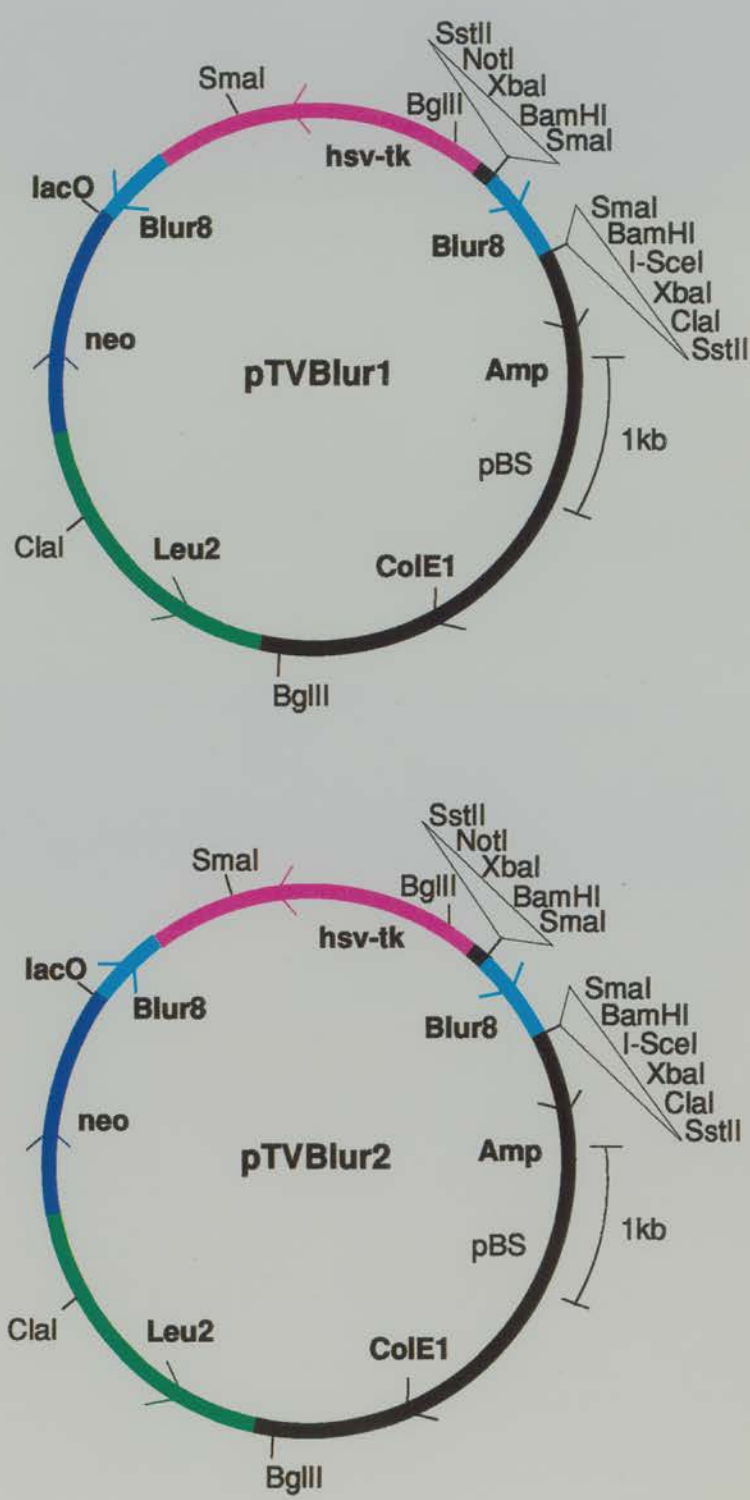
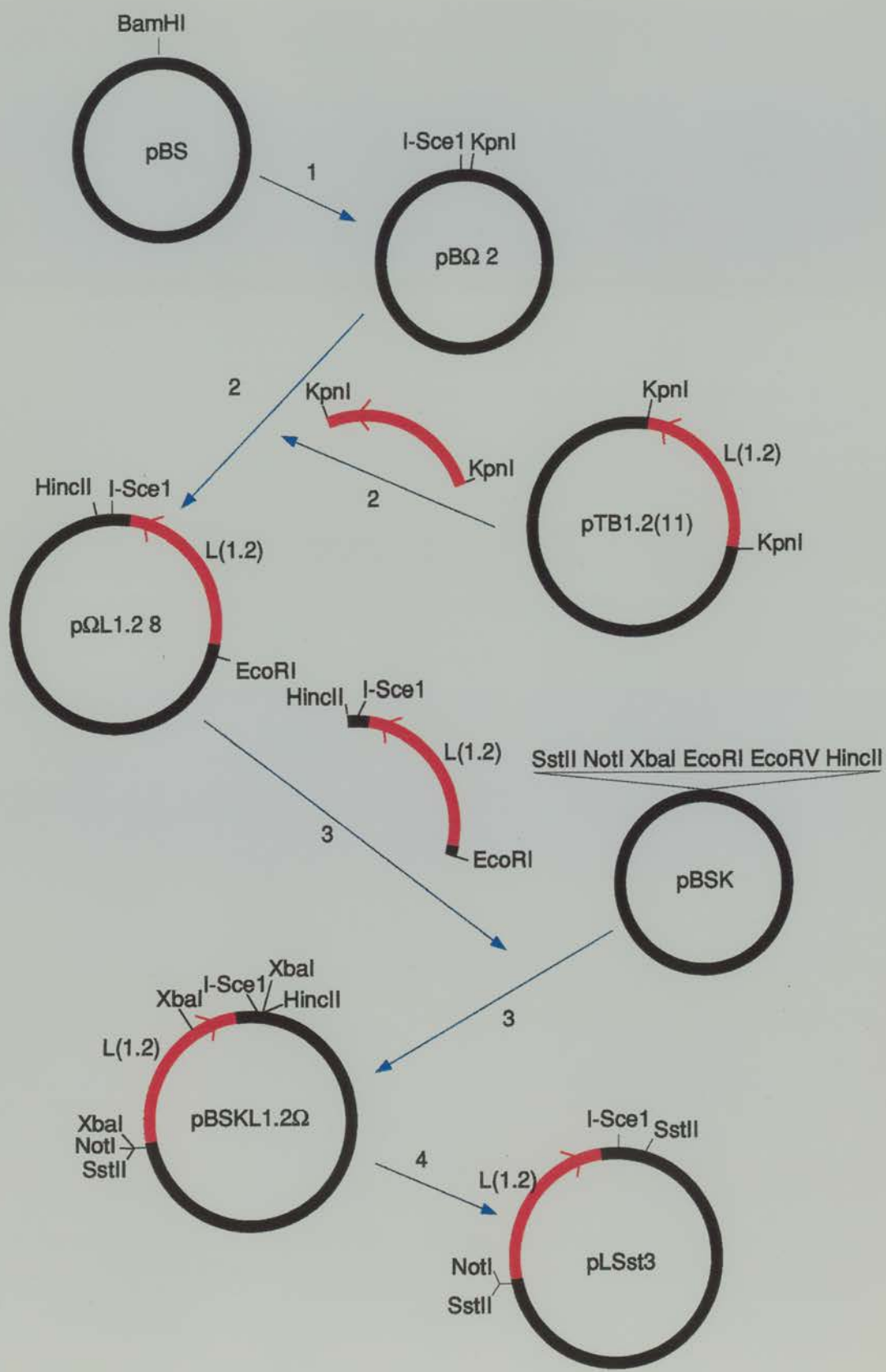


Figure 3.4.0

Construction of pLSst3



colonies were produced, 12 of which were picked and DNA prepared by miniprep. Asp 718 digests revealed which clones contained the required insert; four of these were partially sequenced using the oligo 291 as a primer. p $\Omega$ 1.2 8 was determined as having the insert in the required orientation.

### **3) pBSKL1.2 $\Omega$**

DNA was prepared of a ~1.25kb fragment released by HincII + EcoRV double digestion of p $\Omega$ 1.2 8. This DNA contains the L(1.2) fragment and the I-SceI site of p $\Omega$ 1.2 8. The DNA was ligated to a pBSK vector which had been previously digested with EcoRV and EcoRI and dephosphorylated. The ligation reaction contained ~150ng vector and 60ng of the insert (vector: insert size = 2:1, ratio of ends in the reaction  $\approx$  3:2). The reaction was electroporated into DH10B cells. Following appropriate ligation of sticky and blunt ends, only clones containing the insert in one orientation could be generated. These were identified following XbaI digestion, and one was selected as pBSKL1.2 $\Omega$ .

### **4) pLSst3**

DNA of pBSKL1.2 $\Omega$  was linearised with HincII and then dephosphorylated prior to ligation with phosphorylated SstII linkers, employing the same conditions as those used to create p1 $\Omega$ 2SstII. pLSst3 was thus produced which contains a unique SstII (KspI) fragment of ~1.25kb.

## **3.5 Construction of pLL and PLB**

### **See figure 3.5.0**

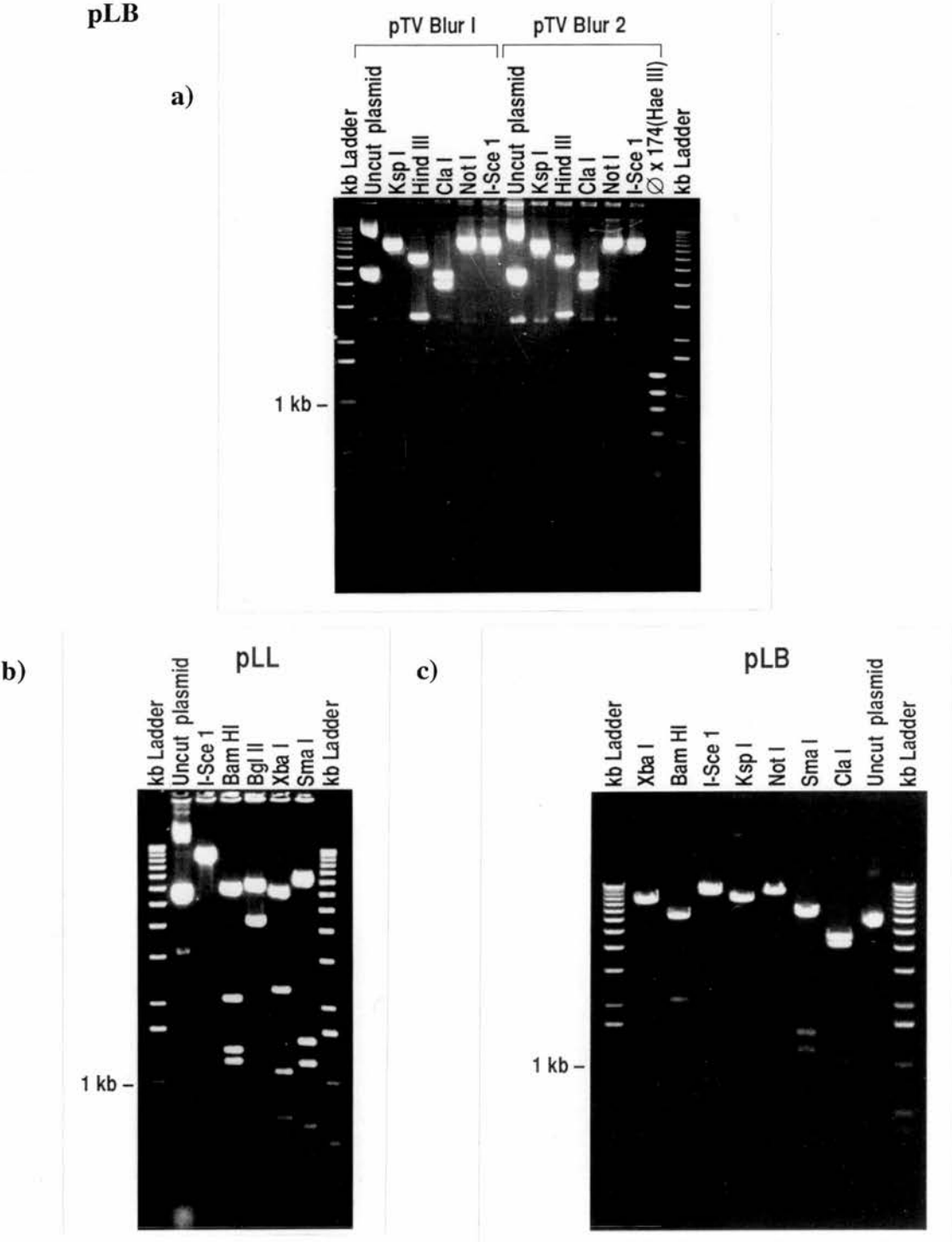
Purified DNA of the 1.25kb insert from pLSst3 was isolated. DNA from pTVL1.3 and pBLOB4XN<sup>-</sup> was linearised by digestion with KspI (SstII) and then dephosphorylated. Prepared vector DNA was ligated to the 1.25kb insert DNA in reactions containing ~20ng pTVL1.3 vector and ~2ng insert (vector : insert size  $\approx$  8:1) and 150ng pBLOBXN<sup>-</sup> with 40ng insert DNA (vector: insert size  $\approx$  6:1). The ligations were transformed into electro-competent DH10B cells and plated out. DNA from colonies produced from both ligations was digested with HindIII. Those clones which had the diagnostic band sizes were identified and were further characterised with digests using I-SceI, NotI, XbaI, SmaI and ClaI (see Fig. 3.5.1). The required clones were named pLL (containing two L1 fragments) and pBL (containing an L1 and a Blur8 fragment). Restriction maps of the completed vectors are shown in Fig. 3.5.2.

**Figure 3.5.0**      Construction of pLL and pLB



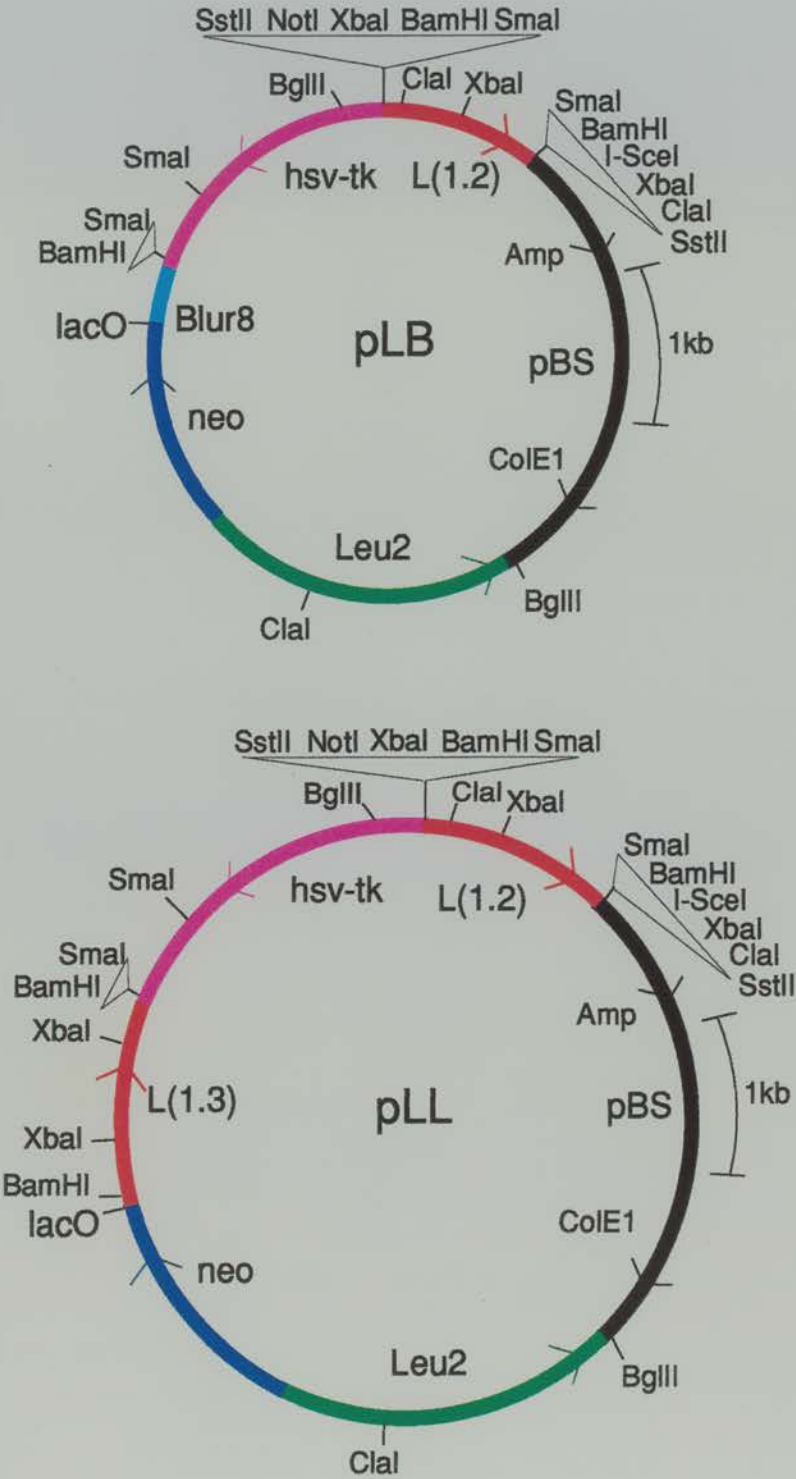


**Figure 3.5.1 Restriction digestion of the vectors pTVBlur1, pTVBlur2, pLL and pLB**



DNA was prepared from each of the completed constructs by the maxiprep protocol (2.3.3). 1.5µg aliquots of each of the vectors were digested with various different restriction enzymes as described in section 2.4.1 (digests with the enzyme ISce-I are described in section 2.4.10). The samples were electrophoresed through 1% agarose gels in 1x TAE at 20v overnight. The gels were stained with 0.5µg/ml ethidium bromide and then photographed. Part (a) shows digests of PTVBlur1 and pTVBlur2; (b) shows pLL; and (c) shows pLB.

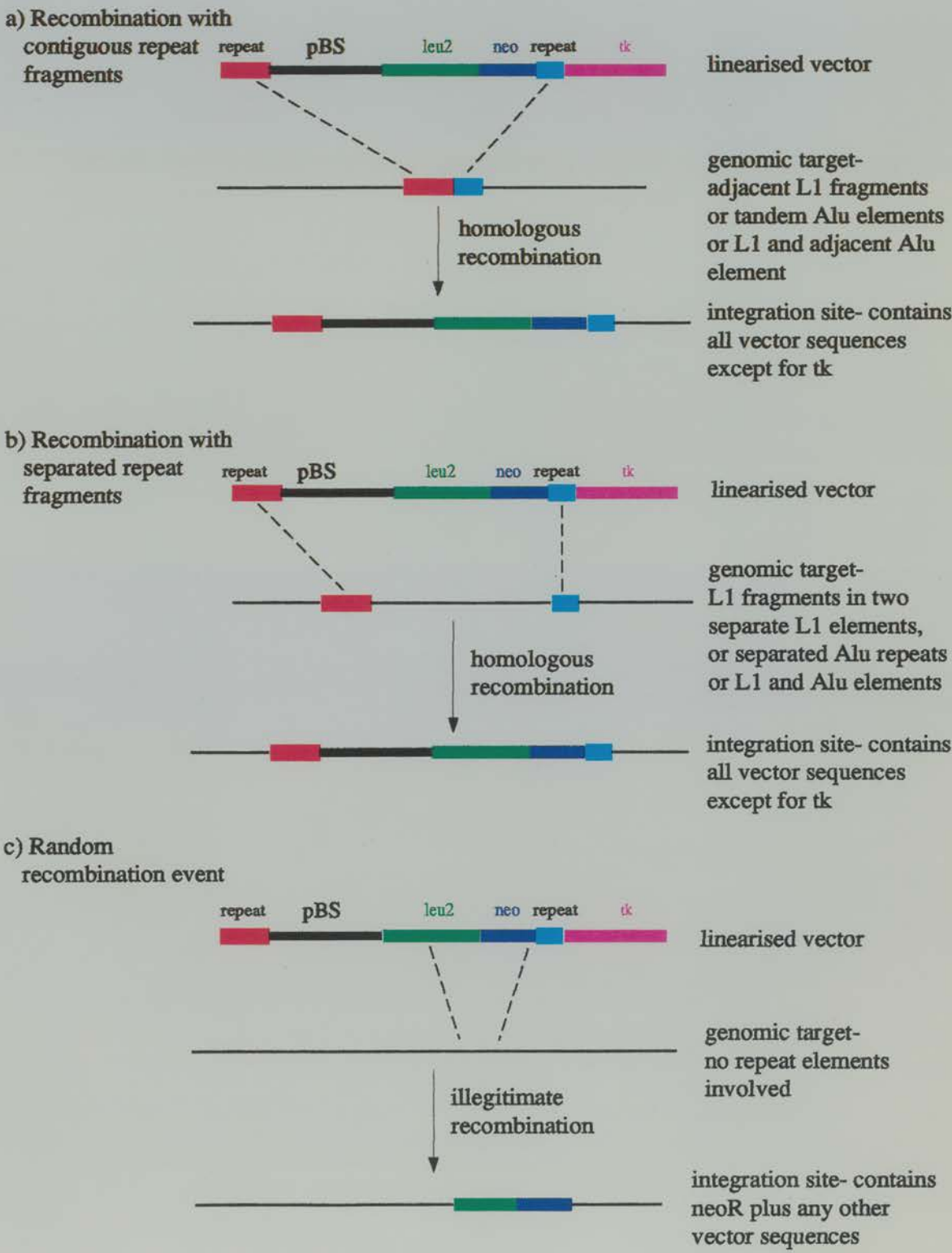
**Figure 3.5.2** Restriction maps of the vectors pLL and pLB (approximately to scale)



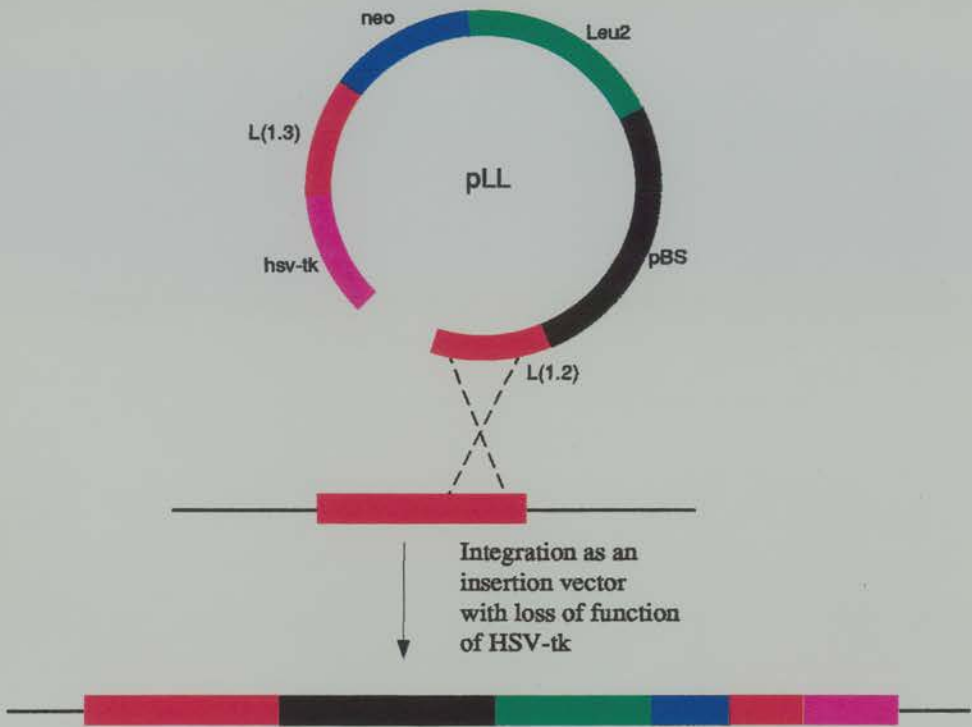
### **3.6 Mechanisms of recombination**

The vectors transfected into the somatic cell hybrid cell lines were designed as replacement-type vectors with positive and negative selectable markers. Depending on the genomic target they may integrate to replace repeat sequences which are adjacent to each other, as is expected of recombination mediated by Line1 elements, shown in Fig. 3.6 (a), or they may replace a larger region of genomic DNA if recombination is with separated repeat elements, as shown in Fig. 3.6 (b). Alternatively they could integrate entirely at random with fortuitous conservation of the positive selectable marker and with loss or inactivation of the negative selectable marker, as shown in Fig. 3.6 (c). They may also integrate as an insertion vector with one or both of the repeat sequences mediating recombination, with loss of function of the negative selection marker, as shown in Fig 3.6 (d). A concatemerisation event may occur followed by either repeat-mediated or illegitimate recombination, as shown in Fig 3.6 (e).

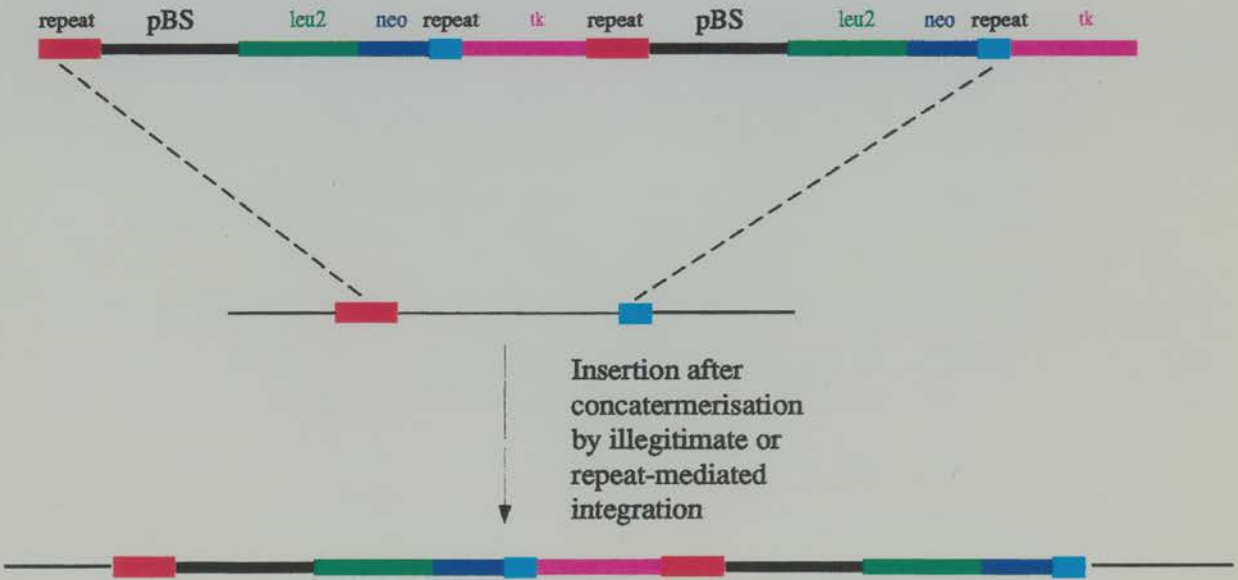
Figure 3.6 Possible recombination mechanisms for vector integration



d) Recombination as an insertion vector



e) Concatemerisation of vector prior to integration





## **CHAPTER 4**

## **4 Transfection of vector constructs and detection of their integration sites by fluorescent *in situ* hybridisation.**

### **4.1 Transfection of vector constructs into J1 c4**

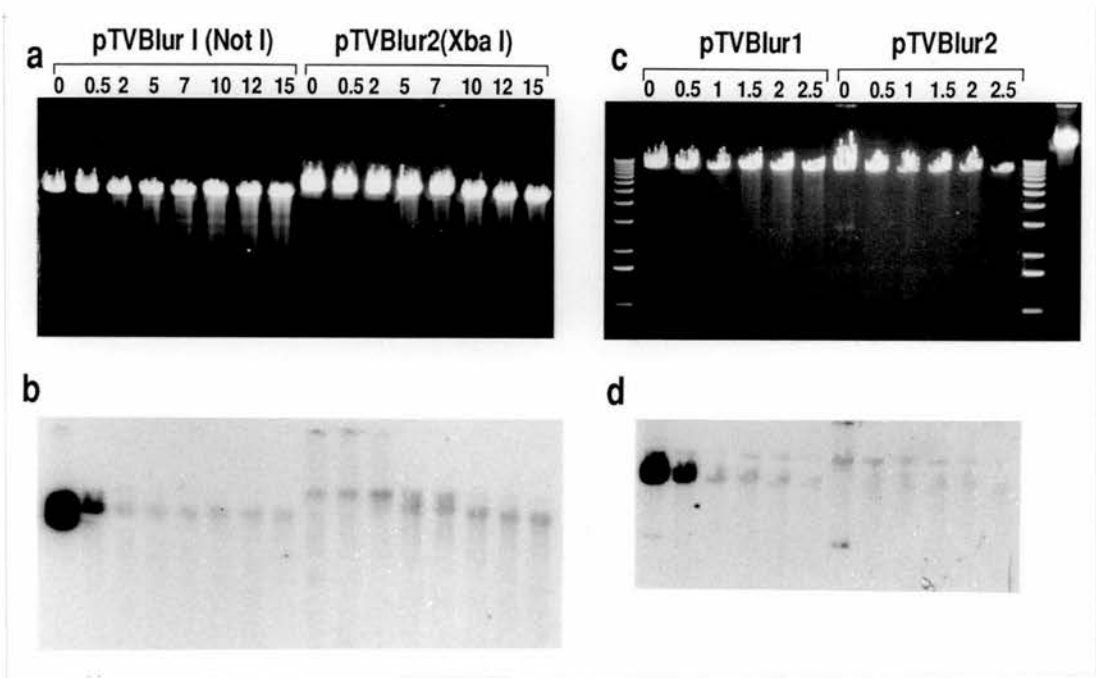
#### **4.1.1 Preparation of vector DNA for transfection**

The aim of the transfection experiments was for the vector constructs to target a whole series of different Line1 and Alu repeat sequences distributed throughout the genome. Prior to introducing the vectors into cells, the constructs were linearised and prepared as described in section 2.5.1. By digesting the ends of the linear vectors with an exonuclease, I hoped to create a set of vectors with different ends and in which possible 'recombinogenic' sequences within the targeting sequences would be exposed. It was proposed that such treatment of the vectors will maximise the chances of recombination occurring with different genomic repeat elements resulting in integration of the markers in many different sites throughout chromosome 11. It also removes a short length of plasmid 'linker' DNA, thus ensuring that the 5' end of the linearised vector lies within the L(1.2) sequence .

Optimisation of Bal31 exonuclease digestion conditions was carried out as described in section 2.5.2. Figure 4.1.1 shows typical results obtained using the vectors pTVBlur1 and pTVBlur2. The time required for the exonuclease to remove 20bp of DNA from pTVBlur1 and pTVBlur2 corresponding to the 'linker' sequence using 3U Bal31 per 100µg DNA was between 0.5 and 2 minutes. To define this time more precisely aliquots were taken at 0.5, 1, 1.5, 2 and 2.5 minutes of incubation. Figure 4.1 shows that 0.5-1 minute of Bal31 digestion under the specified conditions was the time of incubation required to remove the linker DNA. Samples of pTVBlur1 and pTVBlur2 from 0.5 and 1 minute of digestion were pooled and cleaned up (see section 2.5.3) for transfection into cell lines.

The initial set of transfections (experiments 1a -3b) were carried out using the prototype vectors pL1a and pL1b, with pTBPNS3 which contains no repeat DNA as a control. A diagram of these vectors is shown in Fig. 5.1.1. The linearised vectors were treated with 1.8U Bal31 per 100µg DNA. Aliquots from 12.5 and 17.5 min of digestion were pooled and used for transfection into cells. Table 4.1 summarises the processing of vector DNA, the transfection conditions, the selection applied and the clones obtained. Two flasks of cells were transfected with each vector construct; G418 selection alone was applied to one flask and both G418 and gancyclovir selection was applied to the other. This was done in order to estimate the degree of enrichment provided by the negative selection marker tk. The number of clones

**Figure 4.1.1** Bal31 exonuclease digestion of pTVBlur1 and pTVBlur2



50µg of the linearised vectors was digested with 3U/ Bal31 exonuclease/100µg DNA in a total volume of 100µl as described in section 2.5.2. The reactions were stopped after 0, 0.5, 2, 7, 10, 12, and 15 min by adding 3µl EDTA at 4°C. 12µl aliquots of the samples were run out on a 0.8% agarose gel in 1xTAE at 80v for 4hr. The gel was photographed after staining with ethidium bromide, shown in (a). The DNA was transferred to Hybond N by Southern blotting overnight. The blot was prehybridised in oligo hybridisation mix at 48°C for 2hr before hybridising to the 'linker' oligo C308 at 48°C for 4hr (as described in section 2.11). The filter was washed at room temperature in 6x SSC, 0.1% SDS for 10 min and then at 48°C in fresh washing buffer for 20 min before exposure to X-ray film at -70°C for 2hr. The autoradiograph obtained is shown in (b). Part (c) shows a gel run out under the same conditions as described in (a), with samples digested with Bal31 exonuclease for 0, 0.5, 1, 1.5, 2, and 2.5 min. Part (d) shows the result of the hybridisation of a Southern blot of the gel shown in (c) to the 'linker' oligo. The blot and the hybridisation were carried out under the same conditions described for (b).

Table 4.1 Conditions and Results for transfection of J1C4 with vectors L1a, L1b and pTBPNS3

Expt. #.	Vector	Repeats in vector	Enzyme cut with	DNA processing	Amt. of DNA	# cells transfected	Gancyclovir selection	#clones picked	#clones frozen	#clones FISH	Chr. 11 integrations
1a	L1a	1xLine1	Not 1	Bal31-digested, phenol/chloroform extracted, ethanol-precipitated	20µg	2.5x10 <sup>6</sup>	-	9/400	7	8	0
1b	L1a	1xLine1	Not 1		20µg	2.5x10 <sup>6</sup>	+	8/9	3	3	0/1
2a	L1b	2xLine1	Not 1		20µg	2.5x10 <sup>6</sup>	-	6/300	4	4	0
2b	L1b	2xLine1	Not 1		20µg	2.5x10 <sup>6</sup>	+	7/8	4	5	1/3
3a	PNS3	none	Not 1	re-suspended	20µg	2.5x10 <sup>6</sup>	-	0/700	0	n/d	n/d
3b	PNS3	none	Not 1		20µg	2.5x10 <sup>6</sup>	+	12/70	5	n/d	n/d
3c	none	n/a	n/a	n/a	0µg	2.5x10 <sup>6</sup>	-	no survivors	0	n/a	n/a

Table 4.2 Conditions and results for transfection of J1C4 with vectors pTVBlur1, pTVBlur2 and L1b

Expt. #.	Vector	Repeats in vector	Enzyme cut with	DNA processing	DNA quantity	#cells transfected	Gancyclovir selection	#clones picked/ total	#clones frozen	#clones FISH	Chr. 11 integra- tions
4a	pTVB1	2xBlur8 (h-to-h)	Not 1	not de-salted	9µg	1.6x10 <sup>6</sup>	+	0/0	Experiment discarded		
4b	pTVB2	2xBlur8 (h-to-t)	Xba 1		9µg	1.6x10 <sup>6</sup>	+	1/1			
5a	pTVB1	2xBlur8 (h-to-h)	Not1	Bal31- digested, centricon- desalted	30µg	2.0x10 <sup>6</sup>	+	33/40	22	21	1/18
5b	pTVB1	2xBlur8 (h-to h)	Xba 1		30µg	2.0x10 <sup>6</sup>	-	12/25	7	n/d	n/d
6a	pTVB2	2xBlur8 (h-to t)	Not 1		30µg	2.0x10 <sup>6</sup>	+	35/40	27	22	0/19
6b	pTVB2	2xBlur8 (h-to t)	Xba 1		30µg	2.0x10 <sup>6</sup>	-	10/50	4	n/d	n/d
7a	L1b	2xLine1	Not 1	as per expt. 2b	30µg	2.0x10 <sup>6</sup>	+	24/ ~200	24	12	3/8



obtained which survived single selection was between 10- 40 times the number of those surviving double selection. Results were obtained by FISH for the localisation of the site of integration of the vector for some of the doubly-selected clones and the numbers of these are also shown in Table 4.1.

The second set of transfections (Table 4.2, experiments 4a-7a) was with pTVBlur1 and pTVBlur2; vectors containing two Alu repeat units in a head-to-head (h-to-h) and a head-to-tail (h-to-t) configuration, respectively. The vectors were linearised with either XbaI or NotI and digested with Bal31 prior to transfection into the hybrid cells. The number of cells transfected, selection conditions applied and the number of clones obtained are summarised in Table 4.2.

The first experiments (4a) and (4b) failed to produce any viable clones. This was probably due to the presence of high concentrations of salt in the solution of DNA. Liposome-mediated gene transfer depends upon an electrostatic association of DNA with the cationic lipid resulting in a complex with a net negative charge which is then able to interact with cellular membranes. The presence of residual salt in the standard DNA preparation may have resulted in the formation of a lipid/DNA complex which had an altered overall charge and therefore had a sub-optimal interaction with cellular membranes. Alternatively, high salt concentrations could have had direct toxic effects on the Chinese hamster ovary-human hybrid cells. Prior to subsequent transfections with these vectors, the DNA was therefore desalted using a Centricon-30 column (see 2.5.3).

For experiments (5a) to (7a), a 75cm<sup>2</sup> flask of ~90% confluent cells had previously been harvested and divided between 5 x 75cm<sup>2</sup> flasks. Upon reaching ~60% confluence, two flasks of cells were transfected with each of the vectors pTVBlur1 and pTVBlur2 whilst the fifth was contemporaneously transfected with pL1b DNA (as used in the previous experiment), to act as a control. Either single or double selection was applied to the flasks as shown in Table 4.2. The number of clones which survived double selection compared to the number of those surviving single selection was not however significantly different in this experiment. The number of clones on which FISH analysis was performed and for which integration sites were localised is also given in the table.

Two mouse fibroblast cell lines containing human chromosome 17 as their sole human DNA component: GM10321 and GM10498 were also transfected with pTVBlur1 and pTVBlur2 under the same conditions. This was undertaken as part of a collaboration with Maria Stack (from the group of John Boyle, CRC, Christie Hospital, Manchester). Clones surviving double selection were picked and stored.

Table 4.3 Conditions and results for transfection of GM10321 and GM10498 with vectors pTVBlur1 and pTVBlur2

Expt. #. (cell line)	Vector	Repeats in vector	Enzyme cut with	DNA processing	DNA quantity	#cells transfected	Gancyclovir selection	#clones picked	#clones frozen	#clones FISH	Chr. 11 integra- tions
8a (GM10498)	pTVB1	2xBlur8 (h-to-h)	Xba 1	Bal31- digested, centricon de-salted  as above	30µg	2.0x10 <sup>6</sup>	+	24	18	16	0/16
8b (GM10498)	pTVB2	2xBlur8 (h-to-t)	Not 1		30µg	2.0x10 <sup>6</sup>	+	24	9	8	0/5
8c (GM10321)	pTVB1	2xBlur8 (h-to-h)	Xba 1		30µg	2.0x10 <sup>6</sup>	+	24	17	16	1/16
8d (GM10321)	pTVB2	2xBlur8 (h-to-t)	Not 1		30µg	2.0x10 <sup>6</sup>	+	24	3	3	0/3

Table 4.4 Conditions and results for transfection of J1C4 with pLL and pLB

Expt. #. (cell line)	Vector	Repeats in vector	Enzyme cut with	DNA processing	DNA quantity	#cells transfected	Gancyclovir selection	#clones picked	#clones frozen	#clones FISH	Chr. 11 integra- tions
9a (J1c4)	pLB	1xLine1 + 1x Blur8	Not 1	Bal31- digested, 'magi-prep' desalted	20µg	1.6x10 <sup>6</sup>	+	24	24	5	4/5
9b (J1c4)	pLL	2xLine1	Not 1		25µg	2.6x10 <sup>6</sup>	+	24	17	2	1/2

M. Stack performed FISH on metaphases prepared from these clones. Results for these experiments are summarised in Table 4.3.

A third set of transfections into the cell line J1c4 was undertaken with the vectors pLL and pLB. Linearised vector DNA was digested with 3U Bal31 per 100µg DNA and aliquots from 0.5, 0.75, 1 and 1.25 min of reaction time were pooled and used for transfection of J1c4 cells. Conditions and transfection results for these are summarised in Table 4.4, together with the number of FISH results obtained.

## 4.2 Localisation of vector integration sites by fluorescence *in situ* hybridisation (FISH)

### 4.2.1 The sensitivity of FISH

FISH is a powerful technique for gene mapping by direct probe localisation on preparations of chromosomes (outlined in section 1.2.3). However, problems can arise when the length of the probe sequence is short, as is the case for the detection of integration sites of the vector in clones arising from these transfection experiments. When the probe length is short, relatively little antigen can be incorporated into the sequence during the labelling reaction, resulting in small amounts of complexed hapten to detect. Also, in some interactions with its target site a short probe maybe entirely concealed within the tertiary structure of the chromosome and therefore not be recognised by the antibodies required for detection. These problems are ameliorated when the probe can complex with itself to form larger arrays or networks as is the case when longer sequences are used as probes (cosmids for example) (Gerhard *et al.* 1981, Singer *et al.* 1986). Signal from a probe can also be amplified when additional labelled molecules complex with the unbound sequence of the primary probe attached to the target (Persing *et al.* 1989) giving a so called 'Christmas tree' effect.

The plasmid construct pTVLOΩ was used to detect the site of integration of the vectors. It is approximately 8.5kb in size and contains all the sequences present in the targeting vectors apart from the repeat elements. It was expected that in many cases the negative selection marker HSV-tk would be lost at the site of integration enabling some of the labelled fragments of pTVLOΩ to form probe complexes, thus increasing the signal available for detection. The plasmid was labelled with biotin (Bio) or digoxigenin (DIG) by nick translation to give probe fragments labelled to a high specificity. In FISH experiments to detect the site of integration of the 'first generation' vector pL1b a 3.2kb SmaI/BglII fragment containing the *neoR* and *Leu2* genes was used as a probe. pL1b was designed so that only the *neo/Leu2* cassette would be present at the site of integration if recombination had been mediated by the

repeat elements. Perhaps signal would have been improved if the entire plasmid had been used as a probe.

Measures were also taken to enhance the sensitivity of detection of the signal. Confocal laser-scanning microscope systems have been built which can take optical sections along very narrow planes of the sample, allowing the signal-to-noise ratio to be maximised when an image of the sample is produced. Also, a threshold level for signal strength can be set, below which no signal is observed as part of the image, and this allows enhancement of specific signal over that of background. The greater sensitivity of this system means that signal previously undetected or non-resolvable by a human eye assisted by a UV-light microscope can now be detected. Antigens and antibody systems have been improved so that several layers of fluorescent signal can be built up without cross reaction, thus amplifying the signal from the probe. Both these methods for enhancement of low levels of signal were used in the localisation of the sites of integration of the vectors in these experiments.

#### **4.2.2 Alpha satellite DNA as a probe for FISH**

Alpha-satellite (alphoid) DNA is found in the heterochromatin located at the centromere of human chromosomes. It essentially comprises a 171bp repeat unit. These monomeric units make up higher order repeat units that are sufficiently similar within chromosomes and diverged between chromosomes to make the repeats useful as chromosome specific markers (Willard and Waye 1987, Moyzis *et al.* 1987). A human chromosome 11-specific alphoid probe pH53 (Yurov *et al.* 1987) was used together with the vector probe in most of the hybridisation reactions to enable rapid identification of the human chromosome in the metaphases, and to provide a positive control for the reaction. The higher order repeat units of the alphoid DNA are present in hundreds of copies at centromeres so that the chromosome 11-specific probe consisting of 15 monomers will hybridise many times to give a very strong signal at the chromosome 11 centromere. In some of the FISH reactions, however, non-specific hybridisation of signal was observed due to cross-reactivity of the antibodies used to detect DIG and biotin. To resolve this problem in later experiments, pH53 DNA was directly labelled with Rhodamine red or fluorescein by including fluorochrome-bound dUTP in the nick translation reaction (section 2.14.6). As the signal from this probe could be observed without any amplification by tagged antibodies, this simplified the detection procedure (section 2.14.8), as well as reducing non-specific signal.

### 4.2.3 Optimisation of conditions for FISH

In order to increase the signal-to-noise ratio and exploit the sensitivity of the system, the overall background level of non-specific signal has to be as low as possible in FISH preparations. Also, the number of metaphases observed in a hybridisation reaction with one set of probes needs to be maximised in order to obtain enough with clear signal which can be counted.

I have developed a scheme by modifying previously determined conditions for FISH which can be used for the detection of signal from probes hybridising to a sequence as small as 3kb (see section 2.14). The important parameters were determined as being; a) the freshness of the chromosome preparations; b) the number of metaphases per slide; c) the size of the labelled probe fragments; and d) the stringency of the washing conditions. The optimum conditions are outlined below.

a) The chromosome preparations stored in fixative at -20°C had to be less than a month old when they were used to make slides, and the slides had to be between 3 and 9 days old when they were hybridised. This was to ensure that chromosomes were free from proteins and that they had hardened on the slide, a process which reduces chromosome 'fluffing' and improves banding after hybridisation (Trask 1991). At a later stage some success was also achieved using metaphase spreads on slides which had been baked at 80°C for 1 hour and then stored at -20°C in a sealed box containing desiccant. According to Trask (1991) slides can be stored indefinitely in this manner and for future experiments this would be the method of choice.

b) A high density of metaphases per slide was required to enable the scoring of signal from a significant number of chromosomes. This was obtained by synchronising the cell culture using methotrexate to inhibit DNA synthesis and block the cell cycle, and then releasing the cells *en masse* into late S-phase by removing the methotrexate. The length of time the cells were subsequently cultured depended on the individual cell line so cells were harvested at half hourly intervals between 4 and 6 hours after release to obtain an optimum sample. A high percentage of the cells harvested by 'shake off' were mitotic. The protocol is described in more detail in section 2.14.2. Between 10-20 metaphases were scored per slide wherever possible.

c) The size of the probe fragments labelled with hapten needs to be between 100-500bp to minimise non-specific hybridisation. Labelling probes by nick translation allows the size of fragments produced to be controlled by the amount of DNase I added to the reaction. The size was assessed by running a denatured sample of the probe on a 3.5% agarose gel and then examining them either under UV-light after staining with ethidium bromide, or by transferring them to a nitro-cellulose filter and

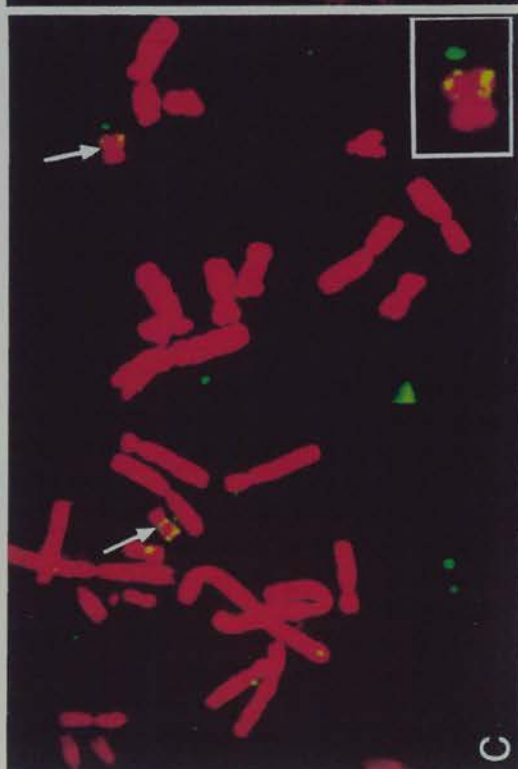
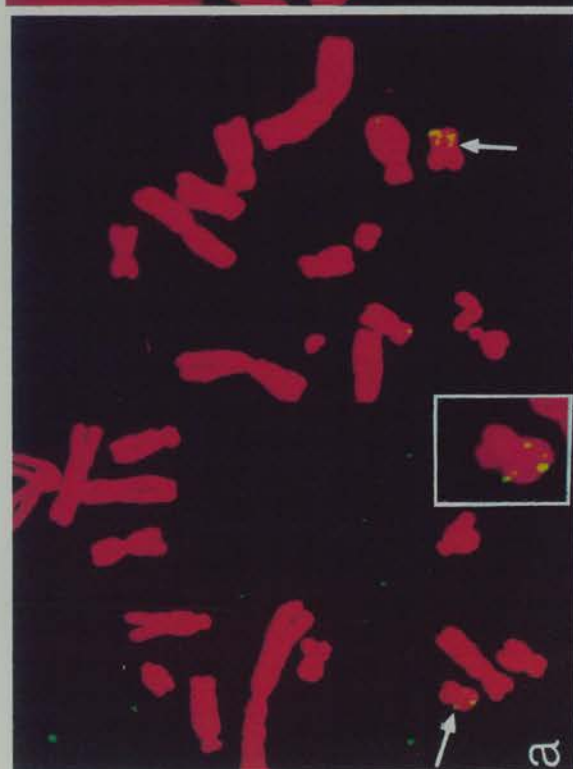


detecting the fragments with alkaline phosphatase-conjugated specific antibody (described in section 2.14.4).

d) The washing conditions had to be stringent enough to remove non-specifically hybridising probe molecules but not so strong as to remove specific signal or to disrupt antibody binding when washing the slides between the layers of detection. By varying the temperature of the washes and the concentration of the buffers, the optimum conditions were determined which are described in section 2.14.8.

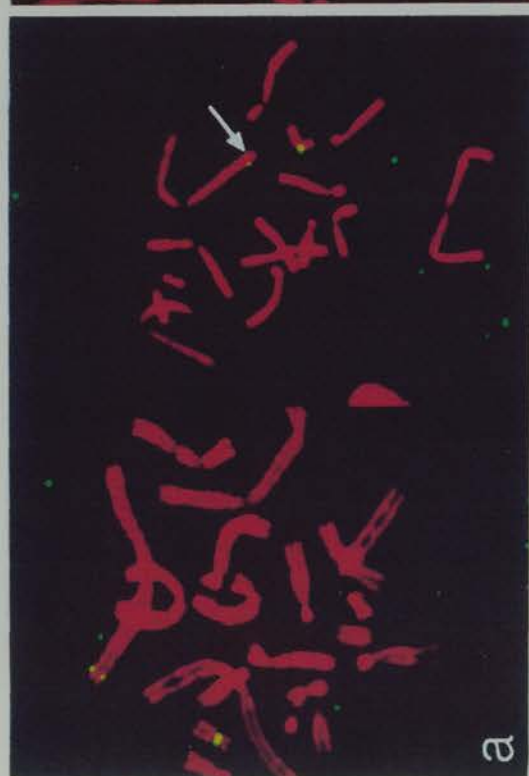
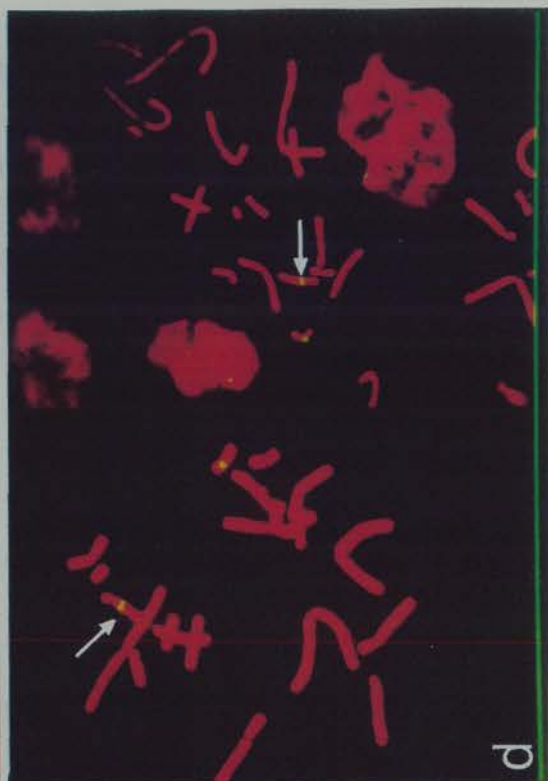
#### **4.2.4 FISH results**

Examples of the results obtained from FISH of vector sequences and chromosome 11-specific probes to clones obtained after transfection of the targeting vectors are given in figures 4.2 onwards. In most cases two images showing different metaphases are given for each clone within each part of the figure. All images except for those in Fig. 4.8 were produced using a BioRad confocal scanning microscope and BioRad 'SOM' software (version 4.62a). Images were photographed at x126 magnification using a Polaroid camera. In some cases the chromosome of interest has been magnified a further 4-fold to enable the signal to be more clearly observed (see insets). Arrows indicate the position of specific signal in some cases.



**Figure 4.2 FISH of clones L1b1(AB), L1b4(AB), L1b12(AB), L1b11(AB), and L1b3(VW)**

Figures (a) to (d) show results of FISH on metaphase chromosomes from clones obtained after transfection of J1c4 with the vector pL1b. All hybridisation reactions were carried out according to section 2.14, unless otherwise stated. A dual-hybridisation was performed in each case using a 3.2 kb *Sma*/BglII fragment from pTBPNSLO which encompasses the *neoR*-*Leu2* cassette, and the IGF2-INS cosmid as a marker for chromosome 11. The *neoR*/*Leu2* fragment was labelled with biotin and detected with three layers of avidin-FITC separated by alternate layers of biotinylated anti-avidin. IGF2-INS was labelled with digoxigenin and detected with layers of FITC-conjugated antibodies, as described in section 2.14.8. The clones shown are (a) L1b1(AB) (two cells), (b) L1b4(AB) (two cells), (c) L1b12(AB) and L1b11(AB), and (d) L1b3(VW) (two cells). Clones L1b1(AB), L1b4(AB), L1b12(AB), and L1b11(AB) arose from a transfection experiment carried out by A.J. Brookes and were resistant to both G418 and gancyclovir. Clone L1b3(VW) arose from my first transfection experiment with L1b where double selection was imposed (expt. 2b, see Table 4.1). All clones show localisation of signal to chromosome 11. The IGF2-INS cosmid gives rise to two adjacent spots of hybridisation at 11p15.5. In all the (AB) clones tested the signal due to hybridisation of the vector probe is localised close to the centromere on 11p.



**Figure 4.3 FISH of clones 1.11, 1.14, 1.17, and 1.18 (pTVBlur1 transfectants)**

This shows the results of FISH on chromosomes from clones arising from the transfection of J1c4 with the vector pTVBlur1 which survived selection with G418 and gancyclovir (see Table 4.2, expt. 5a). Two metaphase spreads are shown for each clone. These are (a) 1.11, (b) 1.14, (c) 1.17, and (d) 1.18. In each hybridisation reaction the probes were biotin-labelled pTVLO $\Omega$  and DIG-labelled pHS53. The vector probe was detected with three layers of avidin-FITC alternated with two layers of biotinylated anti-avidin. pHS53 was detected with a single layer of FITC-conjugated anti-DIG antibodies. In each case the vector probe has hybridised to a Chinese hamster chromosome.



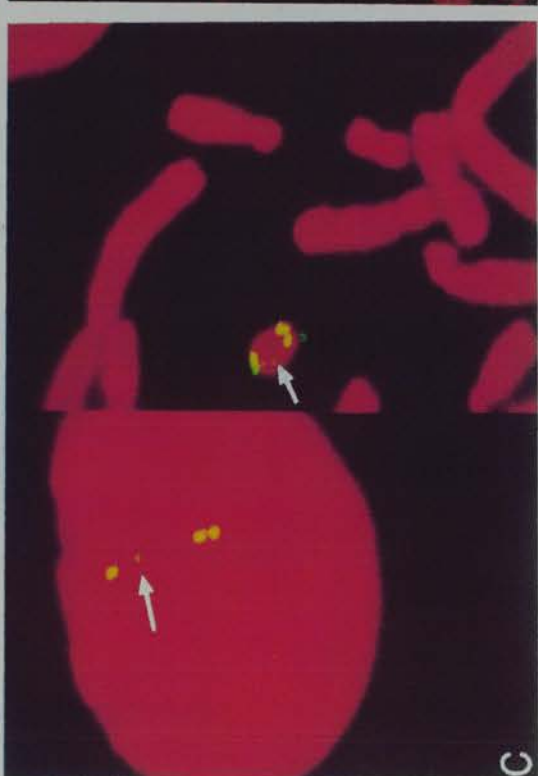
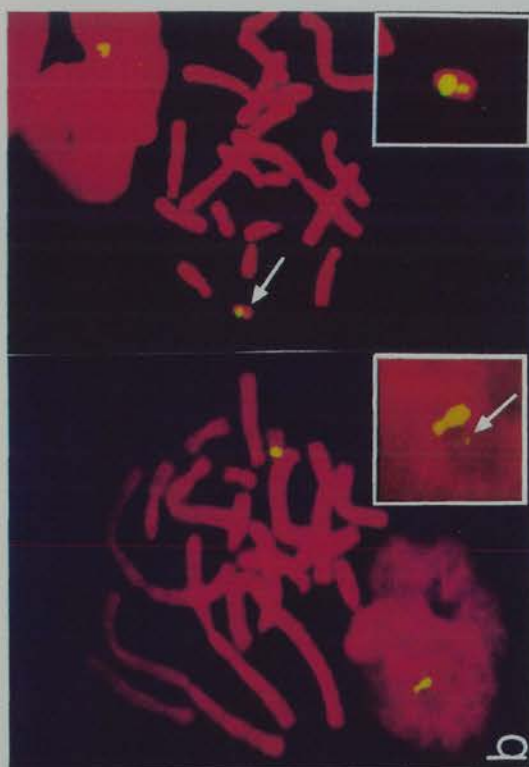


**Figure 4.4 FISH of clones 1.23, 2.37, and 1.22 (pTVBlur1 and 2 transfectants)**

This shows the result of FISH on clones arising from a transfection with vectors pTVBlur1 and pTVBlur2 which survived double selection (expts. 5a and 6a, see Table 4.2). Two metaphases are shown from clones (a) 1.23, and (b) 2.37, on which dual hybridisation of pHS53 and pTVLO $\Omega$  has been performed. Detection of the signal was as described for Fig. 4.5. The vector signal is localised to a Chinese hamster chromosome in both of these clones.

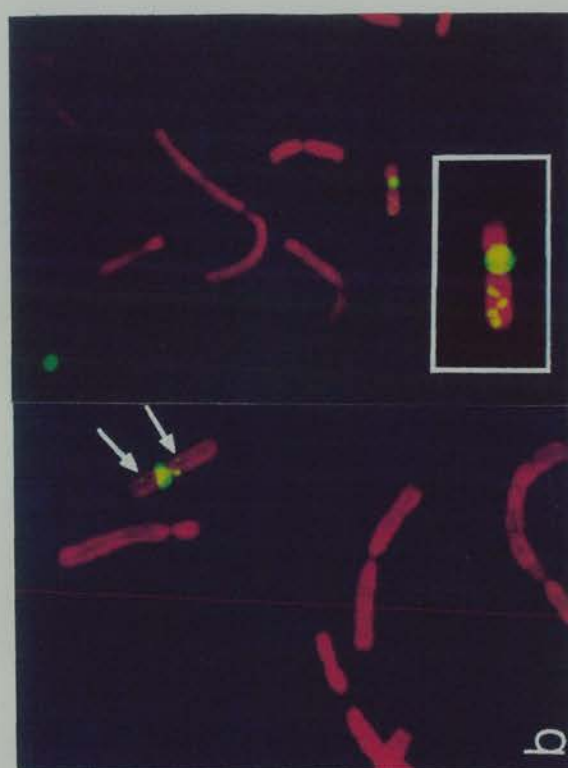
Four separate metaphases with hybridisation are shown of clone 1.22 in (c) and (d) as this was the single clone arising from the transfection with pTVBlur1 and pTVBlur2 in which the specific signal from the vector probe is localised to human chromosome 11, indicating that pTVBlur1 had integrated into the human chromosome in this clone. In the FISH experiment giving the result shown in (c i), a dual hybridisation was performed using biotin-labelled pTVLO $\Omega$  and DIG-labelled pHS53. In the other experiments (c ii) and (d), only the specific probe Bio-pTVLO $\Omega$  was used.

By taking fractional length p-ter (FLpter) measurements on 5 metaphases, a value of 0.668 was calculated for the site of integration. Equivalent values for OMP (olfactory marker protein) and TYR (tyrosinase) are 0.6 and 0.68 respectively (Judy Fantes, pers. comm.). OMP maps to 11q15.5 and TYR maps to 11q14.3 which means that the site of integration of the vector can be assigned approximately to the region of 11q14.2.



#### **Figure 4.5 FISH of clones L1b5, L1b4, L1b8, and L1b2.**

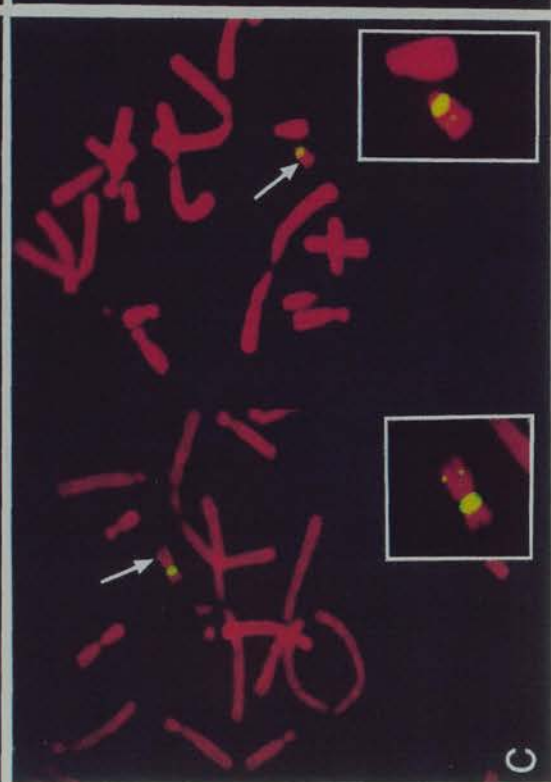
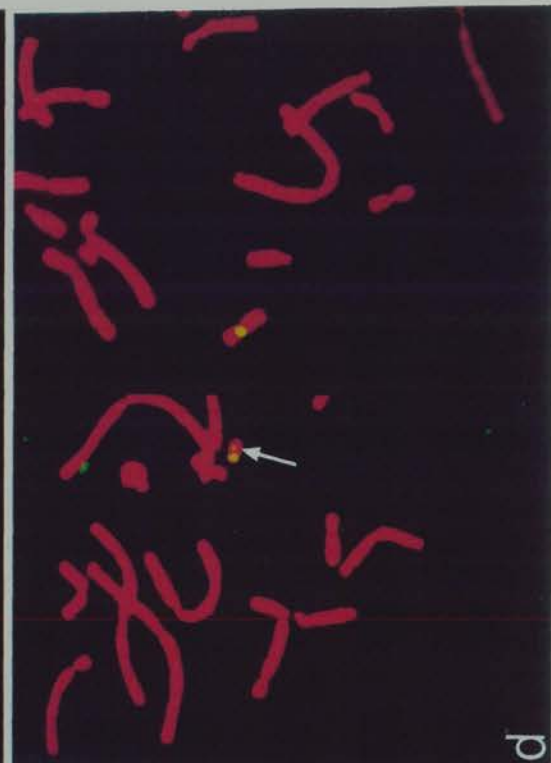
This shows results of FISH on metaphase and interphase chromosomes on clones arising from transfection of J1c4 with vector pL1b (expt. 7a, see Table 4.2). All clones were resistant to G418 and gancyclovir. Two images of different cells are shown for each of the following clones: (a) L1b5, (b) L1b4, (c) L1b8, and (d) L1b2. The interphase and metaphase chromosomes of clones L1b5 and L1b4 were simultaneously hybridised to DIG-labelled pHS53 and biotin-labelled pTVLO $\Omega$ . The signal from pTVLO $\Omega$  was detected with FITC and amplified using alternate layers of avidin-FITC and biotinylated anti avidin. A single layer of FITC-conjugated anti-dig antibodies was used to detect signal from pHS53 (see section 2.14.8). One arm of the human chromosome is clearly shorter than normal in both of these clones. Signal from the vector probe is located towards the centre of the intact chromosome arm in clone L1b4 and is just below the centromere on the intact arm in clone L1b5. The signal from the vector probe can be more clearly distinguished from the strong signal given by the hybridising alphoid probe on the interphase chromosomes for each clone. Clone L1b8 was hybridised to biotin-labelled pTVLO $\Omega$  and DIG-labelled IGF2-INS cosmid. The signal was detected as described for clones L1b5 and L1b4 (above). Two sets of spots on 11pter can be observed (more clearly in the interphase nucleus shown in (c i)), which corresponds to hybridisation of the IGF2/INS probe. The single set of spots at 11qter could correspond to the vector probe, with a second site (seen as a faint spot in the interphase nucleus) more proximal on 11q. The greater intensity of the 11qter signal suggests that this could be due to hybridisation of the cosmid probe as a consequence of a rearrangement of the chromosome 11 in this clone (this is discussed in more detail in section 3 of this chapter). Clone L1b2 was hybridised only with Bio-pTVLO $\Omega$  and detected using alternate layers of avidin-FITC and biotinylated anti-avidin. Specific signal can be seen as two spots on a Chinese hamster chromosome in both metaphases.



### Figure 4.6 FISH of clones LB1, LL2, LB24, and LB5

This shows the results of FISH on some of the clones arising from transfection of J1c4 with the vectors pLB and pLL (expts. 9a and 9b, see table 4.4). All clones selected were resistant to both G418 and to gancyclovir and those clones chosen for analysis by FISH also contained pBS vector sequences as determined by Southern blot analysis (see Chapter 5). Two metaphase spreads are shown from each clone: (a) LB1, (b) LL2, (c) LB24, and (d) LB5. In each reaction, two probes were hybridised to the metaphases; DIG-labelled pTVLO $\Omega$  and biotin-labelled pHS53. Three layers of FITC-conjugated antibodies were used to detect the signal from the vector probe. A single layer of avidin-FITC was all that was required to detect signal from hybridising pHS53. In clone LB1 specific vector signal is localised to the centre of one of the arms of the human chromosome 11; the other arm of the chromosome is absent. In (a i) the termini of the intact arm of two human chromosomes in this clone have become joined together. This effect is observed in an adjacent set of human chromosomes and can also be seen in the interphase of a neighbouring cell (discussed in the text). Two clear sites of pTVLO $\Omega$  hybridisation can be seen on the human chromosome in the metaphase of clone LL2 shown in (b i), one on 11p and one on 11q. Image (b ii) shows two sites on chromosome 11q and possibly a site on 11p. Hybridisation results were obtained with only a limited number of metaphases for this clone, therefore I was unable to confirm the number of sites on the human chromosome detected by the vector. Similarly with clone LB24 (Fig 4.6 c)), although vector signal was localised to the human chromosome, an anomalous number of sites were observed on different metaphases and therefore the integration site could not be precisely identified. It is possible that these clones comprised a mixture of different sub-populations with different integration sites, discussed in Chapter 6. The signal from the vector probe is localised to a Chinese hamster chromosome in clone LB5, shown in (d).







#### **Figure 4.7 FISH of human lymphocyte metaphases and clones LL20 and LB20**

This shows the results of FISH in which pHS53 DNA that had been directly labelled with fluorochrome was used as a probe. Metaphases from human lymphocytes (prepared by Diane Lawson) were hybridised with pHS53 DNA into which FITC-dUTP had been incorporated by nick-translation (described in section 2.14.6) to test out the probe. The results are demonstrated in (a) which shows two metaphases from this reaction. No amplification of signal was required by fluorochrome-conjugated antibodies and all the signal appears to be specific to the centromeres of the chromosomes 11. This probe was therefore used in subsequent dual-hybridisation reactions.

In the following experiments, DIG-labelled pTVLO $\Omega$  was used as the specific probe and was detected using three layers of FITC-conjugated antibodies. Image (b) shows examples of hybridisation to clone LL20; chromosome 11 is clearly labelled at the centromere, and the signal from pTVLO $\Omega$  is localised to a Chinese hamster chromosome. FISH results from four different metaphases of clone LB20 are shown in (c) and (d). Image (c) shows two different metaphases observed on the same slide of the first culturing of the clone. One metaphase shows a chromosome 11 of normal gross appearance with signal localised to mid-11q and the other shows a clearly shortened chromosome 11 with signal still apparent. These two types of cells appeared in approximately equal proportions on the slide. When clone LB20 was re-cultured from frozen stocks, approximately 50% of metaphases observed showed an abnormal chromosome 11 with signal still localised to the centre of the intact arm, and the remaining metaphases had a normal chromosome 11 with no detectable specific signal. These results are discussed in Chapter 6.



#### Figure 4.8 FISH of clone LB15

This shows FISH results with clone LB15. Image (a) is an example of two metaphases from the clone when it was first cultured. The centromere of human chromosome 11 has hybridised with biotin-labelled pHS53, and the specific signal from DIG-labelled pTVLO $\Omega$  can be seen at mid 11q and mid 11p. Detection and amplification of dig was with three layers of FITC-conjugated antibodies and detection of the Bio-pHS53 required a single layer of avidin-FITC. Images (b), (c), (d), (e), and (f) are taken with a CCD (charge coupled device) camera using three filters which has allowed the simultaneous observation of three different colours: FITC, Rhodamine red and DAPI. The vector probe pTVLO $\Omega$  was labelled with DIG and detected using three layers of FITC-conjugated antibodies. The human chromosome 11 alpha-satellite probe pHS53 was directly labelled with Rhodamine red (Fluoro-red, Amersham), and the metaphase chromosomes were stained with DAPI. Images (b) and (c) show that the signal from the vector probe (bright blue dots) is localised to the tip of a long sub telocentric Chinese hamster chromosome at qter and that the human chromosome labelled with pHS53 (red/white signal) is of normal gross morphology. The remaining images, (d), (e) and (f) show examples of metaphases from the same slide where the vector signal is localised to a metacentric Chinese hamster chromosome and the human chromosome in these cell has one arm considerably shortened. These two types of metaphases were present in approximately equal proportions on the slide.

## 4.3 Summary of results for FISH analysis of targeted clones

### 4.3.1 Vector pL1b (1)

Initially, FISH was performed on clones generated by A.J. Brookes using the vector pL1b. The site of integration of the vector was verified for 4/9 doubly-selected clones as being on the short arm of the human chromosome 11, shown in Fig. 4.2 (a), (b) and (c). CHO DNA content of derivatives of the hybrid J1 including J1c4 is approximately  $4 \times 10^9$ bp (Gerhard *et al.* 1987), and human chromosome 11 is approximately  $1.44 \times 10^8$ bp in length (Weissenbach *et al.* 1992). The ratio of human DNA: CHO DNA in J1c4 is therefore approximately 1:28, and consequently 1 clone in 28 analysed was expected to have a vector integration site in the human chromosome if the recombination event was entirely random. To see 4/9 clones with a human chromosome site of integration suggested that the vector was not integrating at random and that it is preferentially recombining with the human chromosome.

I therefore repeated the transfection experiment using pL1b and generated four more clones. The site of integration of the vector was assigned to human chromosome 11 for 1/3 of these clones (L1b3), shown in Fig. 4.2 (d). In this hybridisation reaction the IGF2/INS cosmid appears to be detecting a locus on 11qter as well as two loci on 11p: the specific signal is again located close to the centromere. This unexpected hybridisation pattern of the IGF2/INS cosmid was confirmed by the result of a hybridisation reaction using the cosmid probe alone (data not shown). This suggests that the human chromosome had undergone some sort of rearrangement so that a region from 11p had become located on 11qter. Rearrangements in the human chromosome 11 of J1c4 are not uncommon (P. Byrd unpublished observations). It is therefore proposed that the rearrangement occurred before the integration of vector. The other two clones analysed had spots of specific hybridisation on Chinese hamster chromosomes (see Table 4.1). This result replicated the original observation by A.J. Brookes that a vector containing two L1 elements was able to introduce selectable markers into the single human chromosome in a somatic cell hybrid in preference to the rodent chromosomes.

### 4.3.2 Vectors pTVBlur1 and pTVBlur2

It was proposed that Alu repeat elements might promote a similar effect by mediating recombination with human repeat sequences in preference to heterologous rodent sequences. The vectors pTVBlur1 and pTVBlur2 were derived from the same basic vector as pL1b (pTBPN3) and adapted to incorporate two Alu repeat units. They had also been redesigned so that recombination through the repeat elements would introduce the entire plasmid apart from the negative selection gene HSV-tk. In this

way *Amp* and *ColE1* would be present at the site of integration after repeat-mediated recombination, allowing rescue of the surrounding DNA into bacteria at a later stage. The new design also meant that a longer piece of plasmid DNA would be present at the site of integration and therefore would facilitate its detection by FISH.

Figs. 4.3 and 4.4 show results of FISH on some of the clones arising from transfection experiments with vectors containing two *Alu* repeat sequences. Figs 4.3 (a), (b), (c) and (d) and 4.4 (a) and (b) are a representative selection of the FISH analyses in which 36/37 of the clones showed hybridisation of the vector probe to a Chinese hamster chromosome (see Table 4.2). The site of integration appeared to be different in the majority of the clones indicating that (a) these clones arose from separate integration events, and (b) that recombination was not being directed to a specific Chinese hamster DNA sequence.

In only 1/37 clones generated by this experiment was the vector DNA present in the human chromosome, as detected by the specific vector probe. This result is shown in Fig. 4.4 (c) and (d). It is likely that this clone arose purely by chance and that integration into human DNA was by a random recombination event ( $p > 0.9$ ). Molecular analysis of this clone was undertaken to provide more information on the nature of the recombination event (see Chapter 5).

Transfection of the hybrids GM10321 and GM10498 generated a similar result. In only one of the 40 clones screened by FISH had the vector inserted into a human chromosome (see Table 4.3, FISH results not shown). The ratio of human DNA : rodent DNA in these hybrids is ~1:40 and therefore this event, again, is likely to have occurred by chance and is indicative of a random recombination mechanism.

#### **4.3.3 Vector pL1b (2)**

A second transfection of pL1b into J1c4 cells cultured from the same frozen stock and grown to the same density as those transfected with pTVBlur1 and pTVBlur2 was made at the same time, under the same conditions. (The vector DNA had been prepared slightly differently, see Tables 4.1 and 4.2). This transfection yielded many clones. Of the eight for which FISH results were obtained, three showed that the selectable marker had integrated into the human chromosomes. Some of the results of FISH are shown in Figs 4.5 (a) to (d). Figs. 4.5 (a) and (b) show two images for each of two clones in which the plasmid has integrated close to the centromere of the chromosome 11. In both of these clones one arm of the chromosome is shortened, leaving only a small region of DNA visible on one side of the centromere (this is discussed in section 6.1.2). In the third human-tagged clone, the specific vector signal is located in mid 11q. The human chromosome in this clone also appears to have



undergone some kind of rearrangement as the IGF2/INS probe is detecting a sequence at 11qter: a phenomenon which has been observed previously (see Fig. 4.2 (d)) and detected on a clone using only the IGF2/INS probe.

Fig. 4.5 (d) shows an example of FISH detection of a Chinese hamster integration site for the vector in a clone from the same experiment.

From the results of these experiments it was concluded that L1 elements could indeed direct integration of the vector pL1b to a human chromosome but that Alu sequences in a vector gave no enhancement for human integration events. In the case of the clones arising from integration of pTVBlur1 and pTVBlur2, it appears that illegitimate recombination events are causing the selectable marker to insert into any location within the genome of the hybrid, with no particular locus in human nor Chinese hamster DNA being preferentially targeted. Positive-negative selection (PNS) vectors are designed to enrich for integration events in which recombination has occurred through the specific 'targeting' DNA within the vector. In these experiments it was expected that selection against the HSV-tk gene using gancyclovir would reduce the number of random recombination events observed in surviving clones. The results of the pTVBlur1 + 2 transfections shows that this is not the case. In these vectors, the 5' end of the tk gene is located very close to the end of the linearised construct. It is possible that during exonuclease treatment of the vectors that vital promoter sequences were removed from the tk gene thus rendering it inactive and therefore non-functional as a negative selection marker. This is discussed further in section 6.4.

#### **4.3.4 Vectors pLL and pLB**

Two more targeting vectors were built on the new design and incorporating L1 elements. The vector pLB contains a single 3' L1 fragment and a single Alu element. The vector pLL contains two L1 sequences which are from adjacent KpnI fragments from the 3' end of an L1 element. The vectors were transfected into J1c4 cells. Identification of the position of the site of integration of the selectable marker was made by FISH for seven of the clones arising from this experiment, all of which were known to be positive for vector sequences located between the repeat elements in the constructs, as described in chapter 5 and summarised in Table 5.2. The results are shown in Figs. 4.6, 4.7, and 4.8. Four out of the five clones from the pLB targeting experiment showed an integration site into the human chromosome upon first culturing. One of the two clones from the pLL targeting experiment for which a FISH result was obtained showed a human integration site. Results for the individual clones are described below.



In clone LB1 the integration site was the human chromosome 11, shown in Fig. 4.6(a). However, the chromosome showed a changed morphology with one of the arms considerably shortened, cytogenetically similar to clones arising from a pL1b transfection (see Fig. 4.5 (a) & (b)). It was not immediately clear whether the tagged dicentric chromosomes seen in a few cells and shown in Fig. 4.6 (a i) are a product or a cause of the chromosome shortening. This is discussed further in section 6.1.2. Fig. 4.6 (b) shows a clone from the pLL transfection (LL2), in which the vector has inserted into the human chromosome 11. There appears to be two, perhaps even three sites of integration on the one chromosome indicating that multiple recombination events have occurred in this clone. Multiple integrations of a vector into mammalian DNA is unusual and is discussed further in section 6.4. Another clone from the pLB transfection (LB24), in which the vector appears to have integrated into the human chromosome is shown in Fig. 4.6 (c). This vector also appears to have recombined into multiple sites in the human chromosome. In clone LB5 the vector has integrated into a Chinese hamster chromosome, shown in Fig. 4.6 (d), indicating that vectors containing human L1 elements do not direct recombination solely into human sequences. The Chinese hamster recombination events may be due to an entirely random recombination event or they may have occurred via a targeted event where recombination has occurred with a Chinese hamster sequence that shows a degree of homology with the human L1 element, an L1-Cg repeat for example (Miles and Meuth 1989). A clone with a Chinese hamster integration site was investigated further by molecular analysis (described in Chapter 5).

There was some concern that non-specific cross-reaction of the different antibodies used simultaneously to detect digoxigenin- and biotin-labelled probes may generate spurious results. Biotinylated anti-avidin antibodies are raised in goat, and these may cross-react with anti-dig FITC-conjugated antibodies raised in sheep. An alphoid probe directly labelled with FITC obviated the need for fluorochrome-conjugated antibodies to amplify the signal. This was demonstrated by hybridisation of the probe to human lymphocyte chromosomes, the result for which is shown in Fig. 4.7 (a). This probe was therefore used in subsequent experiments and would be used in all future hybridisations when an alphoid probe is required as a marker. Fig. 4.7 (b) shows a clone from the pLL transfection (LL20) in which signal from the vector probe is located on a Chinese hamster chromosome. This again indicates either a repeat-directed recombination event with a specific Chinese hamster sequence or a totally random recombination event.

Clone LB20 showed that the selectable marker was in the human chromosome 11 when it was first cultured. Hybridisation results from ~20 metaphases showed that

~50% had a full length chromosome 11 and the other 50% had a shortened chromosome 11; both had specific vector signal on the human chromosome. The short chromosome was similar cytologically to that seen in clone LB1 (see Fig. 4.7 (a)). Clone LB20 could consist of a mixed population of clones arising from two separate recombination events. After the clone had been frozen down for storage and then recultured two types of metaphases were still observed. The vector signal was again detected on a shortened chromosome 11 in ~50% of the metaphases, but in the other 50% no specific signal could be detected. It is possible that the original clone contained contaminating cells which did not contain a selectable marker and that this sub-population has expanded upon culturing, despite the maintenance of the culture with G418 for selection for *neoR*. Alternatively, the signal in the second type of metaphase may have been too faint to detect in the second FISH experiment with this clone.

Images (a) to (f) in Fig. 4.9 show results of FISH on clone LB15, in which there is also a mixed population of cells. When initially cultured, FISH analysis of the clone showed that vector sequences were present in at least two different sites on the human chromosome 11, shown in Fig. 4.9 (a). FISH analyses of subsequent cultures of the clone showed two different types of metaphases, present in approximately equal proportions, indicating the presence of two sub-populations. Figs. (b) and (c) show one type in which the chromosome 11 is of normal gross morphology and the vector signal is localised to the tip of the long arm of a sub-metacentric Chinese hamster chromosome. Figs. (d), (e), and (f) show the other type of metaphase observed in which the human chromosome 11 is very much shortened and the vector integration site is localised to a metacentric Chinese hamster chromosome. It is possible that rearrangement of the original clone type occurred so that a part of the human chromosome containing the integrated plasmid has been transferred to a Chinese hamster chromosome to generate one of the secondary clone types observed. Alternatively, the original clone could have consisted of three sub-populations, one with the integration in the human chromosome and two with integration sites in different Chinese hamster chromosomes, one of which also has a shortened human chromosome. It is possible that the clones with Chinese hamster chromosome integration sites had a growth advantage which allowed them to outgrow the other sub clone, surviving the stress of freezing and thawing of the culture more successfully. Molecular analysis of this clone was undertaken in order to shed some light on this phenomenon.

## **CHAPTER 5**

## 5 Molecular analysis of clones generated by transfection experiments

### 5.1. Restriction enzyme analysis

#### 5.1.0 Preface

From the results obtained by FISH it appeared that L1 repeats were able to direct integration of the vector into the human chromosome 11 whereas a vector with only Alu repeats showed a pattern of random recombination, integrating at sites throughout the genome of the hybrid. We wanted to determine whether or not site-specific recombination had actually occurred in the clones in which a L1-repeat-containing vector had been identified as integrating into human DNA. We also wanted to know if integration of a such a vector into CHO DNA was occurring due to recombination with Chinese hamster L1-like repeats (L1-Cg or BamHI repeats) or if it was a random recombination event.

Site-specific gene targeting events can be screened for using PCR or restriction enzyme analysis. In strategies involving an insertion into, or replacement of, an endogenous gene a restriction map and sometimes even the DNA sequence of the target locus has usually been previously determined. This means that homologous recombination events can be identified by specific alterations in the size of a PCR fragment or by changes in the size of restriction fragments detected using a probe from the target locus.

Detection of homologous recombination events in these experiments was not so straight forward as the target for the recombination reaction is present in thousands of interspersed copies. Each target site will therefore have a different restriction map and can not be analysed by locus-specific PCR. Other approaches therefore had to be adopted.

Restriction enzyme analysis using sequences from the vector as probes would reveal some information about the site of integration. Distinct restriction fragment patterns for different clones would tell us if they were unique, i.e. that they had arisen through different integration events. Digestion with an enzyme whose site is not present in the vector itself would allow us to recognise the number of integration sites present in one clone. Hybridisation of such digests with probes from different parts of the vector would also tell us if the sites contained these vector sequences. If a recombination event had been mediated by the repeat elements, then the site of integration would be predicted as containing all the sequences present in the vector which lie between the repeats (see Fig. 3.6). We would therefore expect to detect *Amp* and *ColE1* (Bluescribe sequences), *neoR*, *Leu2* and an *I-SceI* site at the site of integration if the L1 element(s) had directed recombination.



To prove conclusively that an homologous recombination event had caused the vector to integrate, the interacting sequences of the endogenous locus and the vector would have to be characterised. This information could be obtained by sequencing a plasmid rescued from the site of integration which contained outlying genomic DNA as well as the vector sequences. The original sequence of the endogenous target in J1c4 could then be obtained from a cross-hybridising clone identified from a plasmid library constructed from the genome of J1c4. The degree of homology shown between the repeat sequences of the vector and its target could then be assessed. Plasmid rescue of sites of integration would therefore be very useful. Restriction enzyme analysis would allow us to identify fragments which encompassed *Amp*, *ColE1* and genomic target DNA which would be suitable for plasmid rescue.

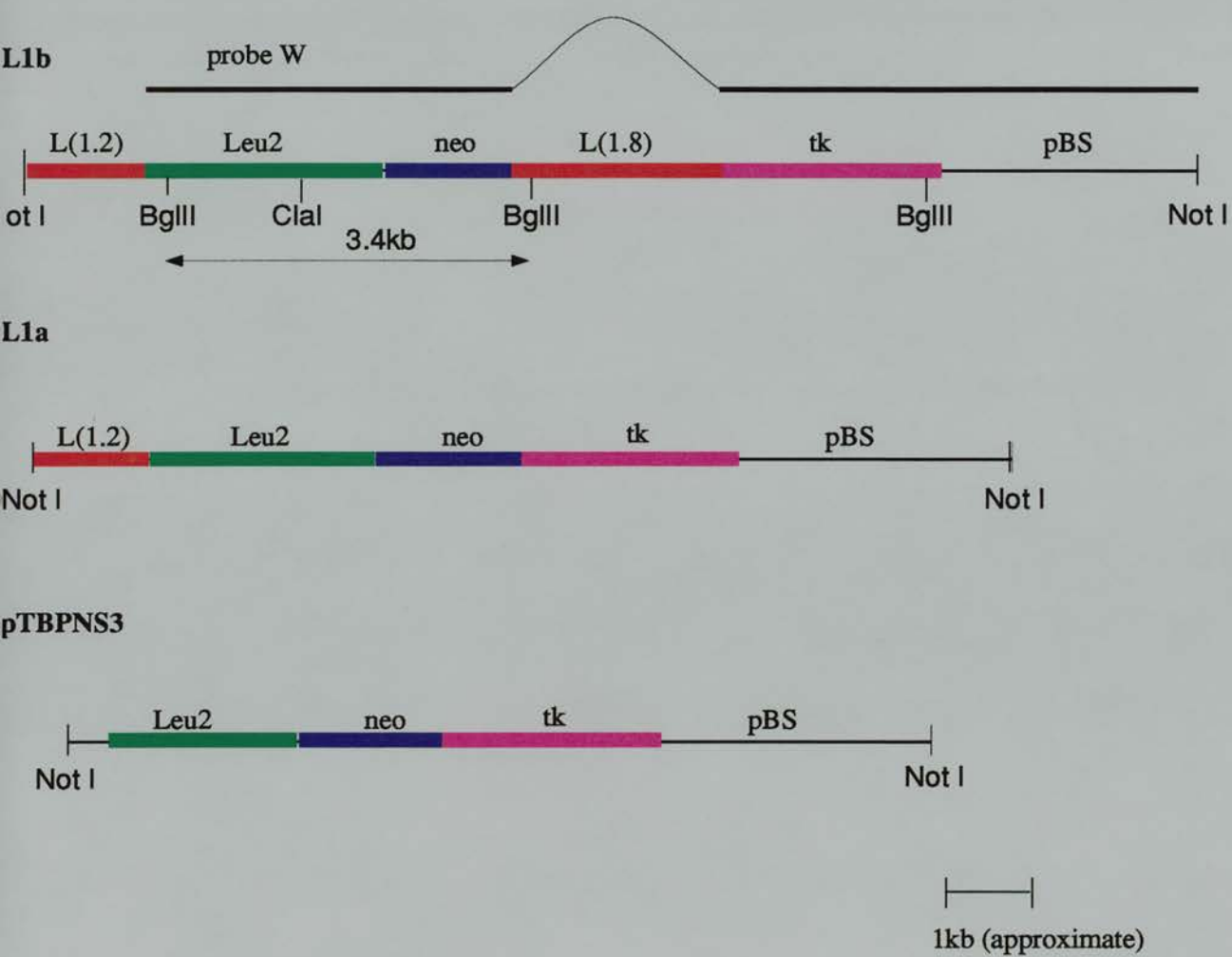
Restriction enzyme analysis was therefore performed on a selection of clones from the different targeting experiments and is described below.

### 5.1.1 Results with the vector L1b

DNA was prepared from G418-resistant clones isolated after the transfection of J1c4 cells with the vector L1b (experiment 7a). DNA from ten of the clones was digested with *Bgl* II, run out on an agarose gel and Southern blotted. The blot was hybridised with the probe pTVLO $\Omega$  (probe W) which is a derivative of pTVLO (see 3.1.2). Probe W contains all of the sequences present in L1b except for the L(1.2) and L(1.8) fragments (see Fig. 5.1.1). The result of the hybridisation is shown in Fig. 5.1.2. The sizes of the cross-hybridising fragments were estimated using the computer program 'gelfit' (written by A.C. Boyd, MRC). A standard curve is plotted by the program using the distances that the marker fragments have moved from the origin on the ethidium bromide-stained gel and the size of the bands detected after hybridisation is then estimated by the distance they have moved from the origin measured on the autorad. All sizes are approximate (usually to within 10%).

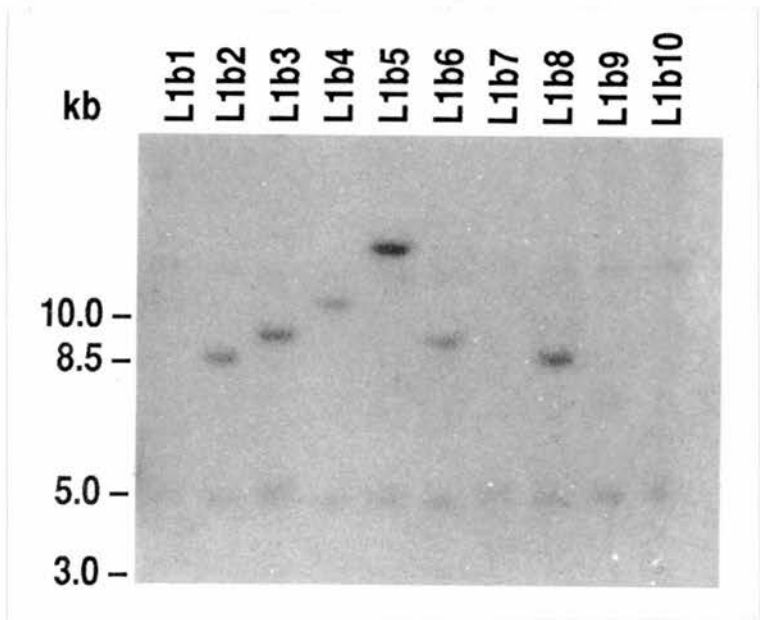
All the clones must contain *neoR* as they are resistant to G418: J1c4 cells not transfected with *neoR* will not survive in the imposed concentration of G418. We would expect to see a fragment of >3.4kb encompassing the *neoR* gene to be detected by the probe if recombination had occurred through the *Line1* fragments to lose the *Bgl*II sites present in the original vector (see Fig 5.1.1). In each of clones 2, 3, 4, 5, 6, and 8, a single fragment of between 8 and 16kb was observed, suggesting that recombination had occurred in the predicted fashion (summarised in Table 5.1). All of the bands are bigger than 3.4kb indicating that one or both of the *Bgl*II sites have been lost during the recombination event. Of the six bands detected, four were of a distinct size, indicating that at least four different clones out of the six analysed have

Figure 5.1.1 Positive/ negative selection targeting vectors (A.J. Brookes 1993)





**Figure 5.1.2 Southern blot hybridisation results of BglII-digests of DNA from G418-resistant clones**



DNA was prepared from clones L1b1-L1b10 as described in section 2.7.1. 10µg DNA was digested with BglII as described in section 2.7.3 and then electrophoresed through a 0.8% agarose gel in 1x TAE at 20v overnight. The DNA was transferred to a Hybond-N filter (Amersham) by Southern blot. The filter was then hybridised with 50ng pTVLOΩ (probe W) radiolabelled with <sup>32</sup>P by random priming (section 2.10). After overnight hybridisation the filter was washed and exposed for four nights to X-ray film before being developed. Conditions for blotting, hybridisation and washing were as described in Chapter 2.

been generated by the transfection experiment. The FISH analysis of clones L1b4, L1b5 and L1b8 (Fig 4.3) shows that the vector has integrated into the human chromosome 11 in each case, and in each clone the site is localised to 11q11-13. The blot analysis demonstrates that these clones are distinct and have therefore arisen independently.

**Table 5.1 Summary of restriction fragment sizes and FISH results for clones arising from transfection of J1c4 with pL1b**

Clone	BglII fragment size (~kb)	FISH result: chromosome with vector integration
L1b1	?	CHO
L1b2	8.5	CHO
L1b3	9.4	CHO
L1b4	11.4	human
L1b5	16.4	human
L1b6	9.1	CHO
L1b7	?	CHO
L1b8	8.5	human

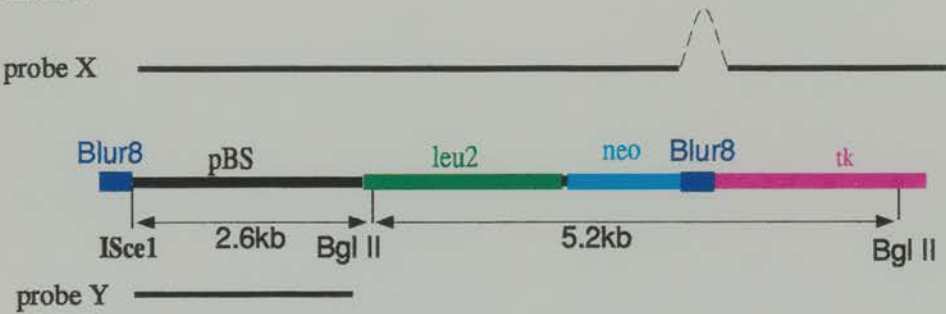
**Summary**

Clones arising from the transfection with pL1b, the 'prototype' vector had BglII restriction fragment sizes compatible with a homologous recombination event mediated by the two L1 fragments in the vector: in none of the clones was the 3.4kb BglII fragment of the vector conserved (see vector diagram in Fig.5.1.1). These data were indicative of a single site of integration in all the clones in which a BglII fragment was detected by probe W. From Table 5.1 it can be seen that both clones in which a CHO chromosome had been tagged with the selectable marker as well as those in which the human chromosome had been tagged have the predicted BglII fragment size.

A homologous recombination event may have introduced the vector into Chinese hamster chromosomes in some of the clones if there was sufficient homology between the L1-H fragments in the vector and certain L1-Cg repeats. Alternatively, the recombination event could have been entirely random with fortuitous loss of either one or both of the BglII sites in the vector in the clones in which the Chinese hamster chromosome was tagged.

Figure 5.2 Map of linearised vector pTVBlur showing the position of probes X and Y

pTVBLUR



There are insufficient data from the molecular analysis to confirm that homologous recombination had occurred to introduce the selectable marker into the human chromosome in 3/8 of the clones, but there is equally no contradictory evidence.

### **5.1.2 Results with the vectors pTVBlur1 and pTVBlur2**

FISH analysis of the 37 clones arising from the transfection of J1c4 with pTVBlur1 and pTVBlur2 showed that the vector had integrated into the human chromosome 11 in only one clone, 1.22, using the vector pTVBlur1. Similarly, clone 321.1.10 was the only one in which the vector had integrated into the human chromosome 17 in the cell line GM10321 with the vector pTVBlur1. In order to analyse the nature of the integration events in more detail, restriction enzyme digests were performed on DNA prepared from these clones. The molecular analysis was compared to that of the clone 1.23 in which pTVBlur1 had integrated into the Chinese hamster DNA after transfection of J1c4.

Digests with SacI, BglII and XbaI were performed, the DNA was then run out on an agarose gel and transferred to a nylon membrane by Southern blotting. The blot was sequentially hybridised with pTVLOXN<sup>-</sup>/BamHI (probe X) and linearised Bluescribe DNA (pBS, probe Y). The plasmid pTVLOXN<sup>-</sup> contains all sequences present in the targeting vectors except for the repeat elements (shown in Fig. 3.1.0.). A diagram of the vector and sequences detected by the probe is given in Fig. 5.2). The hybridisation results are shown in Fig. 5.3. Fragment sizes were estimated using 'gelfit'.

### **Analysis of clones 1.22, 1.23 and 321.1.10**

SacI cuts once within the vector pTVBlur1, in the HSV-thymidine kinase gene. A homologous recombination event with concomitant loss of tk was therefore expected to give only a single fragment encompassing the site of integration that would be detected with both probes X and Y. The only XbaI site in the vector lies adjacent to the NotI site at which the vector was linearised prior to Bal31 digestion and transfection. It was expected, therefore, that the XbaI site would have been removed by the action of the nuclease and that an XbaI fragment detected by either of the probes would encompass the entire site of integration. BglII cuts within the vector twice, once within the HSV-tk gene, and once between *Leu2* and pBS sequences. After a repeat-mediated integration event it was therefore expected that two BglII fragments would be detected by probe X and only one of these would also hybridise to probe Y.

**Figure 5.3 Restriction enzyme analysis of selected clones arising from transfection with pTVBlur1**

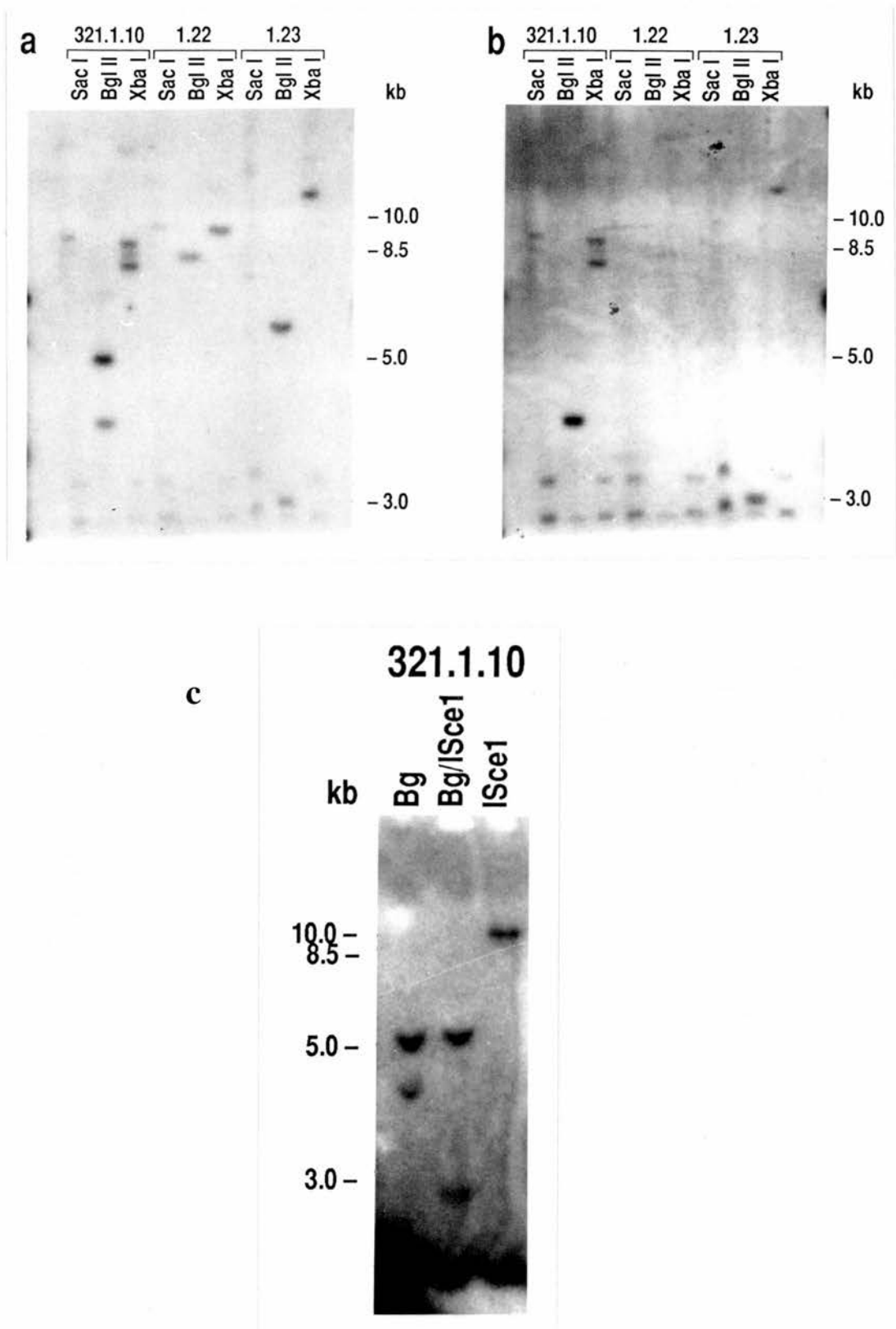
**(a) and (b)**

DNA was prepared from the clones 321.1.10, 1.22, and 1.23 as previously described (2.7.1). 10µg DNA from each clone was digested with SacI, BglII, or XbaI as described in section 2.7.3. The samples were electrophoresed through a 0.8% agarose gel in 1xTAE at 30v overnight. The gel was then blotted and the filter hybridised to radiolabelled pTVLOXN<sup>-</sup> (probe X), as previously described (see sections 2.10.1 and 2.11). The filter was washed and exposed to X-ray film for 4 nights. The developed autoradiograph is shown in (a). The blot was stripped (see section 2.11.5) and after checking that residual signal had been removed was then hybridised to 50ng linearised Bluescribe DNA radiolabelled by random priming (probe Y). Part (b) shows the result of the hybridisation.

**(c)**

DNA was prepared from clone 321.1.10 as described before. The DNA was digested with I-SceI or BglII or both as described in section 2.7.3. Samples digested with both enzymes were first digested with I-SceI and the sample ethanol precipitated before digesting with BglII. The samples were then electrophoresed through a 1% agarose gel in 1x TAE at 20v overnight. The DNA was transferred to a Hybond-N filter by Southern blotting, and the filter was then hybridised to radiolabelled pTVLOXN<sup>-</sup> (probe X), as previously described (see sections 2.10.1 and 2.11). The filter was washed and exposed to X-ray film for 4 nights. The developed autoradiograph is shown in (c).

**Figure 5.3 Restriction enzyme analysis of selected clones arising from transfection with pTVBlur1**





In clone 321.1.10, probe X detects SacI fragments of ~9.5kb and >20kb (faint hybridisation), only the smaller of which hybridises to probe Y (Figs. 5.3 a and b). The BglII digest of the clone gives strong bands of ~5.5kb and ~4.3kb and a faintly hybridising band of ~6.6kb with probe X but probe Y only detects the ~4.3kb fragment. With an XbaI digest fragments of ~8kb, ~8.5kb, ~9kb and a faint band >20kb hybridise with probe X, whereas probe Y hybridises only to the 8, 8.5 and 9kb bands.

An I-SceI digest of clone 321.1.10 produced a fragment of ~10kb detected by probe X (pTVLOXN<sup>-</sup>), shown in Fig. 5.3(c), and a BglII/I-SceI double digest produced the band pattern we would expect from an intact plasmid, indicating conservation of the 2.6kb pBS fragment of the vector. These results suggest that the vector has concatemerised before integration so that more than one intact units of vector DNA are present at the site of integration. This would give rise to the strongly hybridising bands, and the more faintly hybridising fragments produced by all three enzymes could be those containing a smaller length of vector DNA which flank the tandem array and also contain genomic DNA. A concatemerisation event may conserve more than one XbaI site which would explain the multiple XbaI fragments detected by the probes. Another possibility is that the multiple XbaI fragments arose due to incomplete digestion of the genomic DNA. A small rearrangement with an insertion of an extra ~1.3kb between the vector molecules before concatemerisation could explain why the I-SceI fragment and SacI and XbaI fragments are slightly larger than unit length.

An alternative explanation for the observed Southern blot hybridisation results is that the sample was contaminated with plasmid DNA which produces the strongly hybridising bands and that the faintly hybridising bands correspond to the actual site of integration, though this would not explain the ~10kb I-SceI fragment.

The fainter bands are not detected by probe Y (pBS), however, with or without a concatemerisation event, (shown in Fig 5.3(b)) and therefore it is unlikely that the Alu elements have mediated the integration event in this clone.

From Fig. 5.3(a) it can be seen that in clone 1.22, probe X (pTVLOXN<sup>-</sup>) detects a single ~7.7kb BglII fragment which is not detected by probe Y (pBS) (b). The ~9.2kb XbaI and ~9.3kb SacI bands detected by probe X are also not detected by probe Y and this suggests that there is a single site of insertion of the vector which contains one BglII site, the *neoR* gene, and no bluescribe sequences. There is another BglII site ~ 7.7kb away, which must lie outside of the vector sequences. Considering that the clone is resistant to gancyclovir it was thought likely that the site within the HSV-tk gene had been lost during the recombination process, although this is not

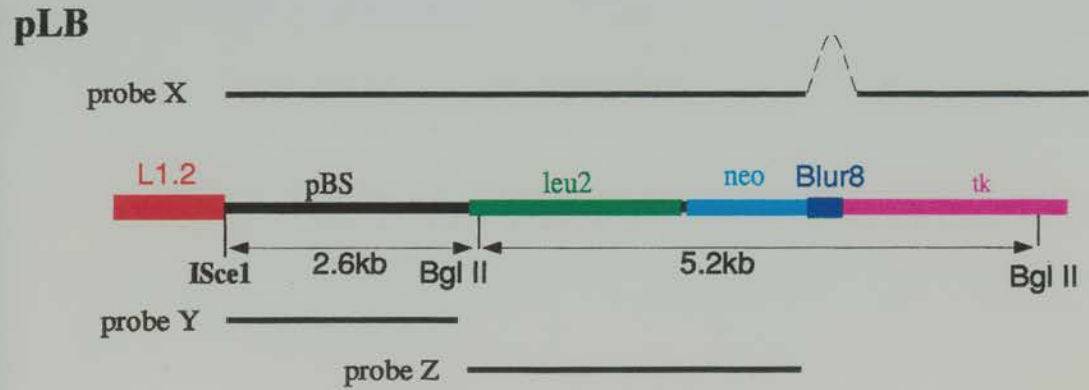
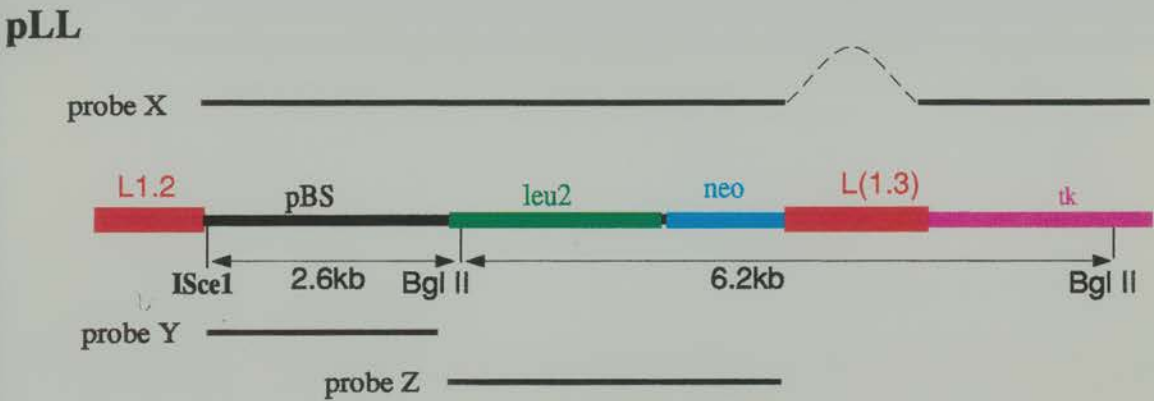
necessarily the case, as discussed later. Hybridisation of the blot with a probe containing just the *Leu2* sequence would have confirmed if the other BglII site from the vector has been preserved. Clone 1.22 therefore has a single site of insertion into the human chromosome which contains only a minimal amount of vector DNA, which suggests that the integration event was random. This is supported by the fact that this was the only clone out of 36 analysed that showed a human integration event; a frequency of no more than 1 in 28 was predicted for a purely random integration event.

Clone 1.23 was one of the clones with a Chinese hamster integration site arising from the same transfection experiment. Probe X (pTVLOXN<sup>-</sup>) detects two BglII fragments in clone 1.23, one of ~5.5kb and one of 3kb, shown in Fig. 5.3 (a). Only the 3kb fragment is detected by probe Y (pBS) as shown in Fig. 5.2 (b). Probe X also detects a ~12.6 XbaI fragment. This data fits with there being a single site of insertion of this vector in which the 5.5kb *neo/Leu2/tk* fragment has remained intact. A smaller BglII fragment which encompasses pBS sequence is also present, and the whole site of insertion is contained within a 12.6kb XbaI fragment. It is possible that the tk gene is present in this clone indicating that recombination is unlikely to have occurred through the repeat sequences by homologous recombination. However, the 5.5kb BglII band could have arisen fortuitously due to a proximal BglII site within genomic DNA. This clone may therefore have arisen through repeat-mediated integration.

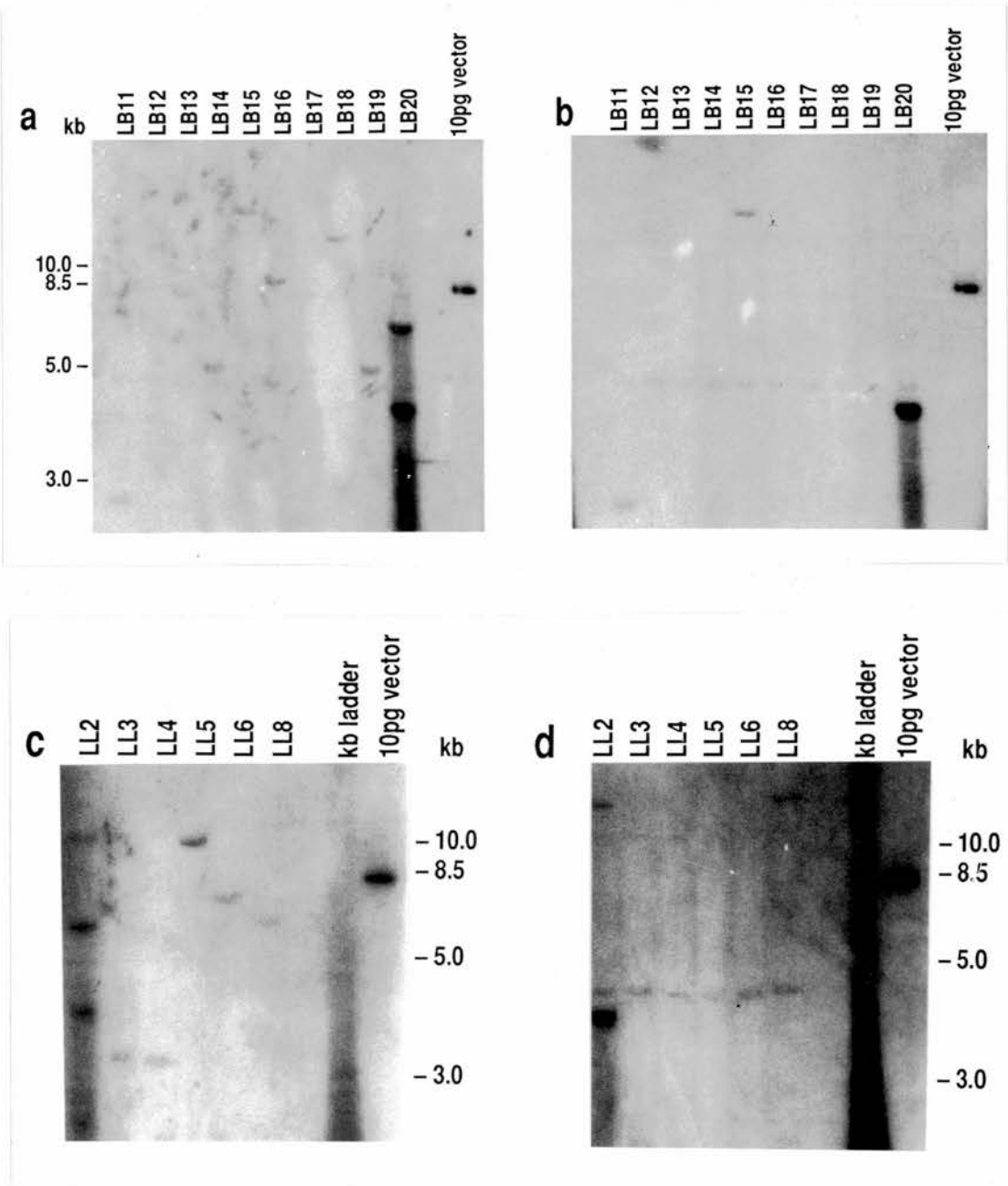
### Summary

The clones with a tagged human chromosome (1.22 and 321.1.10) did not have restriction patterns compatible with a homologous recombination event, agreeing with the results from FISH analysis of 37 clones indicating that recombination was occurring at random. A clone in which the vector had integrated into a CHO chromosome analysed in the same way (1.23) did have a restriction fragment pattern compatible with a homologous recombination event. It is possible that the Alu sequences have promoted an illegitimate recombination event in clone 1.23, causing the vector to integrate at random but mediated by the repeat elements. Alternatively, it is possible that the integration was entirely random and that pBS sequences and BglII sites have been fortuitously conserved.

Figure 5.4 Map of linearised vectors pLL and pLB showing position of probes X, Y, and Z



**Figure 5.5** Southern blot analysis of clones arising from transfections with the vectors pLL and PLB



DNA from each of the clones was prepared and digested with BglII as previously described. Samples were electrophoresed overnight through 0.8% agarose gels at 25v in 1x TAE. The gels were blotted and the filters then hybridised with radiolabelled Bluescribe DNA (probe Y). The washed filters were exposed to X-ray film for 4 nights. The results are shown in (b) and (d). The same blots were then stripped and hybridised with radiolabelled pTVLOXN- (probe X). The results are shown in (a) and (c).

### 5.1.3 Results with the vectors pLL and pLB

DNA was prepared from all G418-resistant clones generated by transfection of J1c4 with the vectors pLL and pLB. The DNA was digested with BglII and run out on gels which were then transferred to nylon membranes by Southern blotting. The blots were then sequentially hybridised first with probe Y (pBS) and then with probe X (pTVLOXN<sup>-</sup>). Diagrams of the linearised vectors and sequences detected by the probes are shown in Fig. 5.4. A sample of the results is shown in Fig. 5.5.

As shown in Fig. 5.4, BglII cuts twice within the vector, once within the HSV tk gene and once within the *Leu2* gene. If a recombination event has occurred in a straightforward fashion, mediated by the repeat sequences in the vector, we would expect to see loss of the 5.2 or 6.2kb BglII bands. Instead, two novel BglII fragments should be seen, one >4kb detected by probes X and Z and one >2kb detected by probes X and Y. The pattern of the BglII digests should also reveal whether or not the clones are distinct.

From Fig. 5.5 (a) and (c) it can be seen that most clones gave BglII fragments which hybridised to probe X. This was expected as they all contain the *neoR* gene which allowed their selection on G418. It was expected that BglII would yield fragment(s) of the size range detected on the blots as this enzyme recognises a site which is present once every ~3kb in the human genome. Some clones (LB1, LB5, LB11, LB16, LB17, LB20, LB24, LL2, LL9, and LL20) showed at least two bands cross hybridising with probe X; this was to be expected if integration has occurred by recombination through the repeat sequences. Some clones contained BglII fragments which hybridised to both probe X and probe Y (pBS) shown in Fig 5.5. (a), (b), (c), and (d), indicating that integration had occurred so as to introduce most of the vector sequences into these clones. These data are summarised in Table 5.2, located at the end of this chapter.

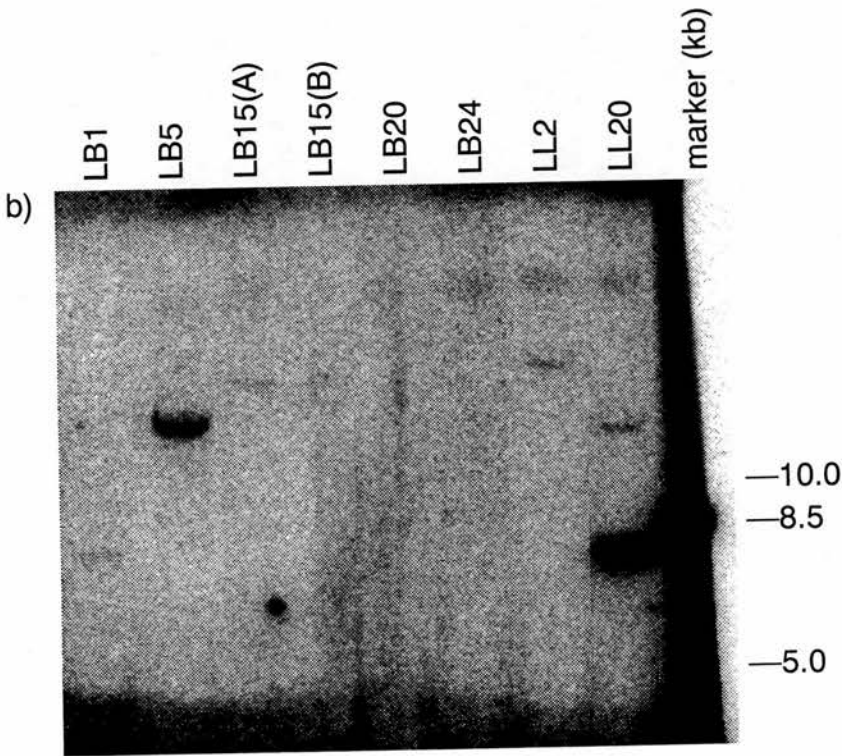
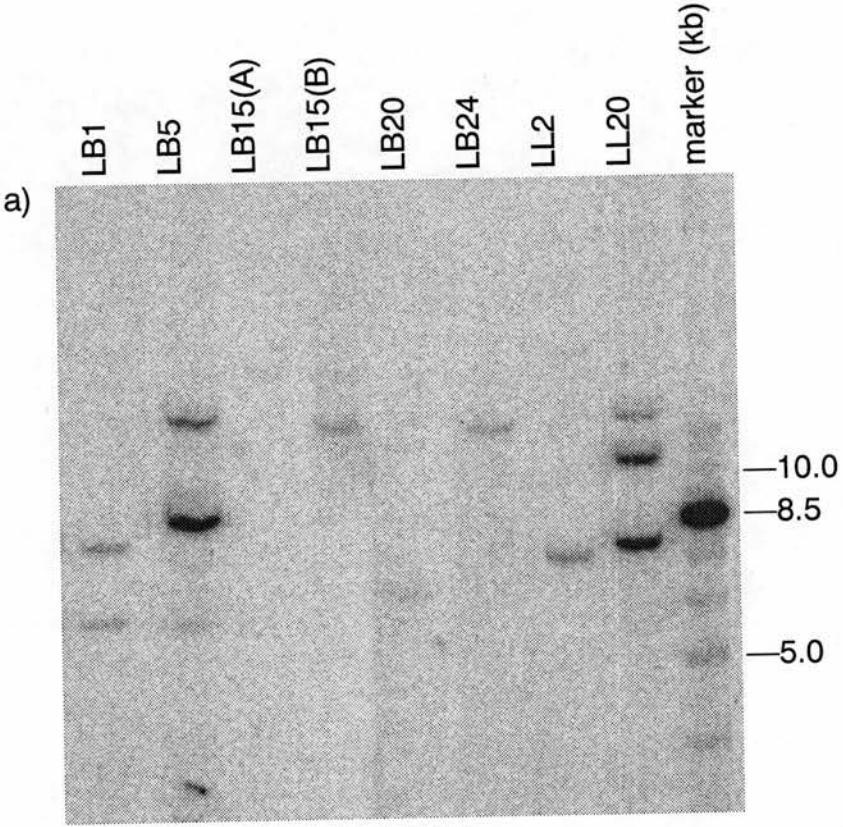
A subset of the clones was analysed in more detail by digesting DNA from the clones with NsiI and with BglII/NsiI in a double-digest. NsiI is a restriction enzyme whose recognition site is absent from both pLL and pLB vectors. It should therefore cleave outside of the site of insertion, providing information on the number of insertion sites present in the clones. Hybridisation of NsiI digests of the clones with probe Z (a 3.2kb SmaI/BglII fragment from pTBPNSLO containing *neoR*/*Leu2*, shown in Fig. 5.4) gives a band pattern corresponding to the number of integrations of the selectable marker in the clones. It was observed that all NsiI fragments detected by probe X (pTVLOXN<sup>-</sup>) were also detected by probe Z which suggested that all sites of integration of the vector seen on this blot contained the *neoR* selectable marker. A hybridisation result with probe Z is shown in Fig. 5.6(a). Probe Z

### **Figure 5.6 NsiI Restriction analysis of selected clones**

DNA from clones LB1, LB5, LB15, LB20, LB24, LL2, and LL20 was prepared and 10µg was digested with NsiI according to protocols described in Chapter 2. The samples were electrophoresed through a 0.8% agarose gel in 1x TBE at 70v for 4.5h. The gel was blotted and the filter hybridised with radiolabelled pTVLOXN<sup>-</sup>(probe X) under conditions described in sections 2.9 and 2.11. The washed filter was exposed to a PhosphorImage screen overnight. The screen was then scanned (as described in section 2.11.4) to give the image shown in (a). After stripping, the filter was exposed to a PhosphorImage screen overnight to check that signal had been completely removed. The filter was then hybridised to radiolabelled Bluescribe DNA (probe Y). The PhosphorImage result is shown in (b).



**Figure 5.6** NsiI restriction analysis of selected clones



(*neoR/Leu2*) and probe Y (pBS) should detect mutually exclusive BglII fragments as a BglII site separates these two sequences in the vector. NsiI/BglII double digests and BglII single digests of the clones were run out on gels which were then blotted and sequentially hybridised with probes Y, X and Z. The results are shown in Figs. 5.5 and 5.7. The sizes of the detected fragments were estimated using 'gelfit'.

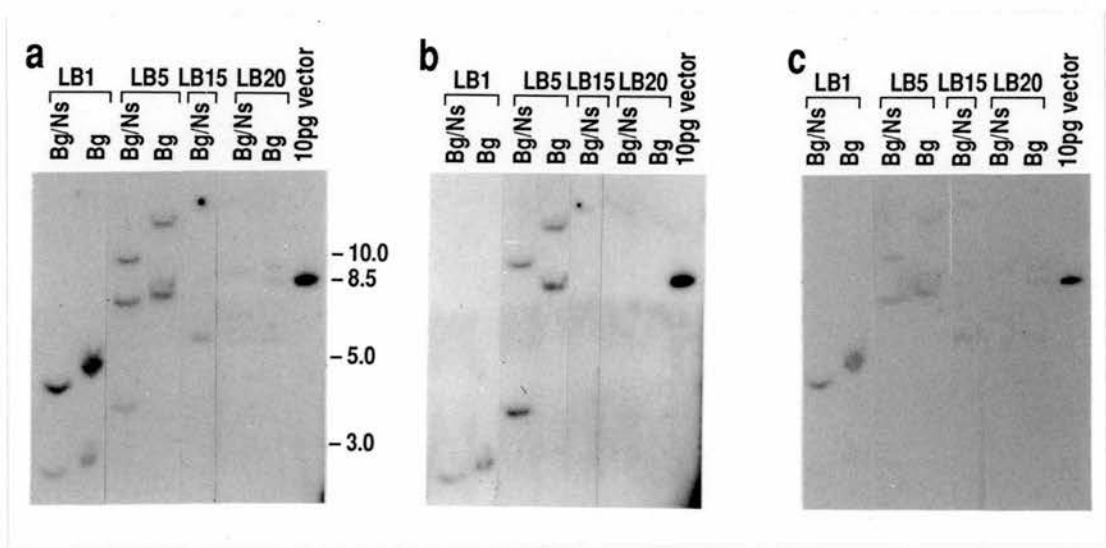
#### **Clone LB1**

From Fig. 5.6(a), probe Z (*neoR/Leu2*) detects two NsiI fragments of ~7.6kb and ~5.7kb in LB1 DNA. Only the ~7.6kb fragment shows clear hybridisation with probe Y (pBS) (b). The NsiI/BglII double digests of LB1 DNA give two fragments which are detected by probes X (pTVLOXN<sup>-</sup>) and Z (~4.4kb and ~2.5kb), shown in Fig. 5.7(a) and (c), only the smaller of which is detected by probe Y. The BglII-digested DNA was slightly overloaded, altering the motility of fragments in this track which consequently can not be directly compared to the other fragments on the gel. However, it can be seen that again two fragments are detected by probes X and Z one of ~4.8kb and one of ~2.8kb, the smaller of which shows only faint hybridisation to probe Z. The 2.8kb fragment is also detected by probe Y (Fig. 5.7(b)). It was unexpected that a BglII fragment should be common to both probes Y and Z. Perhaps the BglII site separating the pBS and *neoR* regions of the vector has been lost during the integration event, or there has been a complicated concatemerisation event preceding integration. We can conclude that there are two sites of integration in LB1, one of which contains bluescribe sequences. The '4.8kb' BglII and the 4.4kb BglII/NsiI fragments detected by probes X and Z hybridise with significantly stronger signal than do the smaller fragments in these tracks, suggesting that perhaps there has been a duplication event at the site of integration of the vector around the *neoR* gene. However, without further restriction digest analysis, a map of neither of the two sites can be constructed.

#### **Clone LB5**

Probe Z (*neoR/Leu2*) detects two NsiI fragments in LB5 of ~14kb and 7.8kb and possibly a third of ~5.7kb (shown in Fig.5.6(a)). This suggests that there are at least two sites of integration of the vector in this clone. Only one of these contains pBS vector sequences, as only the larger ~14kb fragment is detected by probe Y (pBS) (shown in Fig.5.6(b)). The BglII and BglII/NsiI double digests are consistent with there being two sites of integration in this clone, one containing pBS sequences. Fig. 5.7 shows that BglII/NsiI fragments of 11kb and 7.5kb are detected with probe Z (c), fragments of 11kb and 3.8kb with probe Y (b) and fragments of 11, 7.5 and 3.8kb detected by probe X (pTVLOXN<sup>-</sup>) (a). The BglII digest gave a similar pattern of bands, with three being detected by probe X: ~16.5kb, ~9kb and ~8.3kb. Both

**Figure 5.7** Restriction analysis of clones LB1, LB5, LB15, and LB20.



DNA was prepared from the selected clones, and 10µg of each sample was digested either with BglII alone or BglII with NsiI in a double digest. Samples were electrophoresed through a 0.8% agarose gel in 1x TBE at 70v for 5hr. The DNA was transferred to a Hybond-N filter by Southern blotting (see section 2.9.1). The blot was then sequentially hybridised with Bluescribe DNA (probe Y), PTVLOXN- (probe X), and then *neoR/Leu2* fragment (probe Z). The filter was stripped and checked before being re-used each time. The hybridisation results shown are (a) probe X, (b) probe Y, and (c) probe Z.

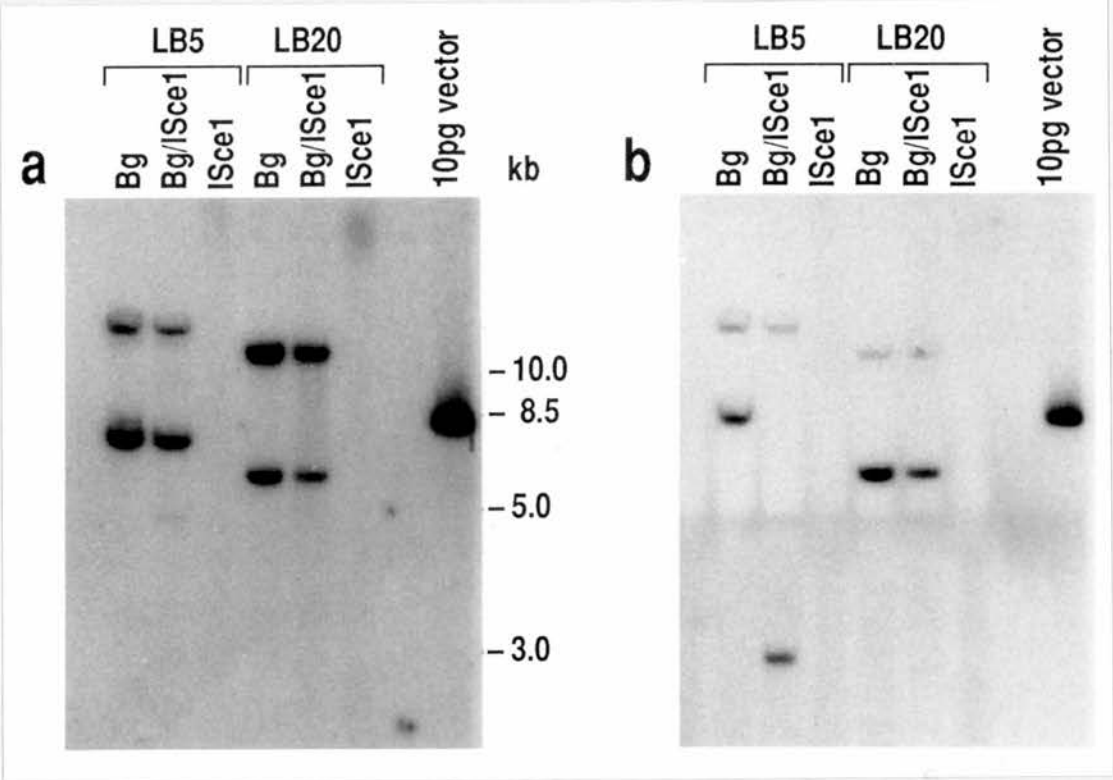
probes Y and Z hybridise with the 16.5 kb band, probe Y also detects the 9kb band and probe Z detects the 8.3kb band. A BglIII/I-SceI digest blotted and hybridised with probe Y shows that the 9kb BglIII fragment contains a functional I-SceI site which divides the band, giving a BglIII/I-SceI fragment of 2.6kb detected by probe Y, shown in Fig.5.8. This vector fragment has therefore been preserved at the site of integration and suggests that the adjoining L1 fragment mediated the integration event. A concatemerisation involving the other end of the vector would explain why the 16.5kb BglIII fragment is detected by all three probes. A possible map of the integration site has been constructed and is shown in Fig. 5.9.

### **LB15**

From the FISH analysis of clone LB15 comparing metaphases from cells harvested from different passages of the clone it appeared that this clone could have undergone chromosomal rearrangement, induced or selected by freezing and thawing of the sample. DNA prepared from two different passages of the clone was therefore digested and compared; sample A is from the original culture, sample B is from a later culture. From Fig. 5.6 (a), it can be seen that NsiI digests of the two samples of LB15 DNA hybridised to probe Z give unpredicted band patterns. LB15(A) DNA gives a single band of ~16kb whereas LB15(B) gives two bands, one of ~16kb and one of 11.7kb. The ~16kb band is also detected by probe Y (Fig. 5.5 (b) in both samples, albeit rather faintly in LB15(B). BglIII/NsiI double digests of the two samples give different band patterns: probes X and Y detect a fragment of ~5.5kb common to both samples and a fragment of >12kb in LB15(A) (data not shown) with a larger fragment >>12kb detected in LB15(B) by all three probes, shown in Fig. 5.7 (a), (b) and (c).

These data suggest that some rearrangement has occurred within the clone around the site of integration during the culture of the clone. Alternatively, the culture might actually contain a mixed population of clones with the selectable markers in different locations in the different clones, and that the proportions of these species within the culture has changed during their growth. FISH was performed on metaphases prepared from LB15 at the two different times of culturing. Initially, there appeared to be two sites of integration in the human chromosome in those metaphases where hybridisation could be detected. After the sample had been frozen down and then re-amplified FISH showed two types of metaphase, one with a shortened human chromosome 11 and one with an integration site into a Chinese hamster chromosome (see Fig. 4.8). The short chromosome 11 could have arisen following rearrangement after integration of the vector, and perhaps part of the human chromosome has been transferred onto a Chinese hamster chromosome. This

Figure 5.8 I-SceI restriction analysis of clones LB5 and LB20.



10 $\mu$ g of DNA from clones LB5 and LB20 was digested with BglIII (as described previously), or with I-SceI (as described in section 2.7.3), or both. Samples digested with both enzymes were digested with I-SceI first and the sample ethanol-precipitated before digesting with BglIII. The samples were then electrophoresed through a 0.8% agarose gel in 1x TBE at 70v for 5.5hr. The DNA was transferred to Hybond-N by Southern blotting and the filter was then hybridised sequentially to Bluescribe DNA (probe Y) shown in (b), and then to the *neoR/Leu2* fragment (probe Z) shown in (a).



could be tested by hybridisation of the most recent passage of clone LB15 with total human DNA so that such a translocation would be observed.

Alternatively, the short chromosome 11 could have arisen prior to transfection by the vector, and is being co-cultured with a clone containing the selectable marker. LB15 could consist of a mixed population of three clones, one with a shortened human chromosome, one with the *neoR* gene in a Chinese hamster chromosome, and one with the marker in the human chromosome. The sub-population with the Chinese hamster integration site might have a growth advantage over those with the human integration, conferred for example, by higher expression of the *neoR* gene due to the presence of two copies or its location in a more highly expressed region. This sub-population could therefore have outgrown the human chromosome 11-targeted sub-population during prolonged culture or periods of stress induced by freezing or thawing of the culture. Chromosome shortening and stability is discussed further in section 6.1.2.

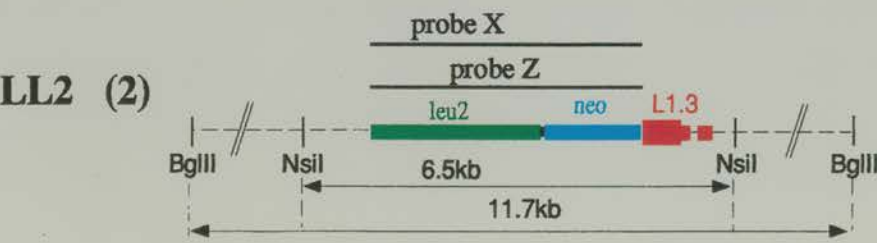
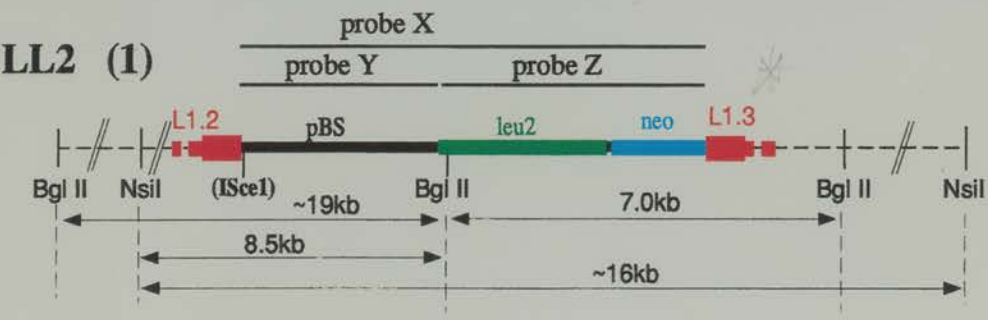
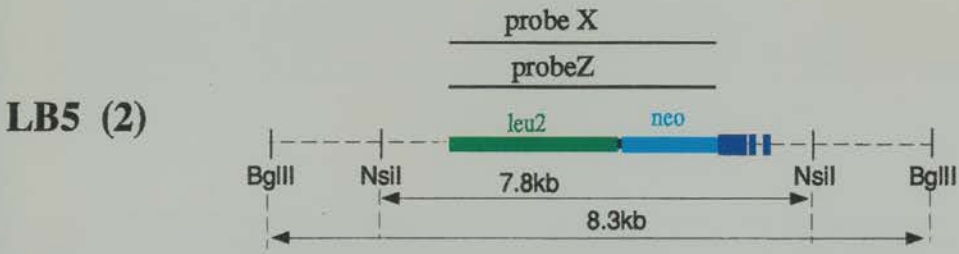
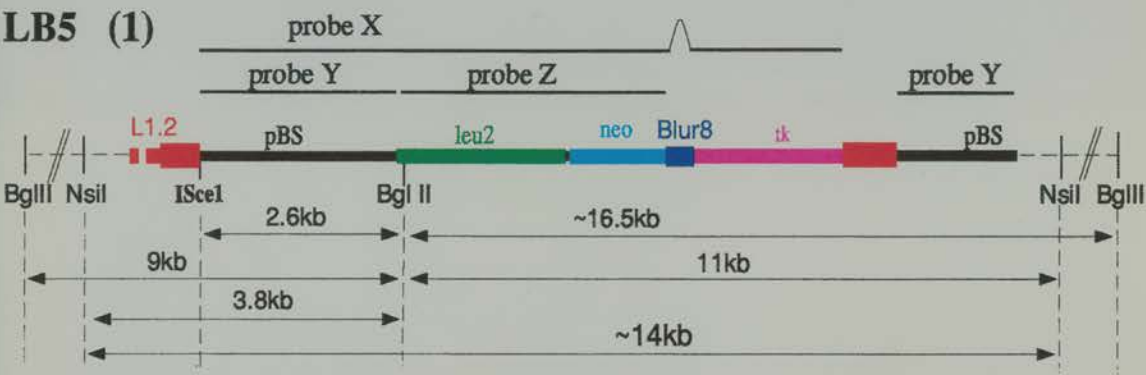
### **Clone LB20**

The restriction enzyme analysis of clone LB20 is shown in Figs. 5.6, 5.7 and 5.8. Fig. 5.6 shows that *Nsi*I fragments of ~10kb and ~6.3kb are detected by probe Z (*neoR/Leu2*) (a), but neither hybridise to probe Y (pBS) (b). The *Bgl*III and *Bgl*III/*Nsi*I digests shown in Fig. 5.7 give three fragments which hybridise with both probes X (pTVLOXN<sup>-</sup>) (a), and Z (c), but again none are detected by probe Y (b). *I-Sce*I digests show that no functional recognition site for *I-Sce*I is present in this clone (Fig. 5.8). This suggests that there are two sites of integration of the vector in this clone and that only a minimal amount of vector DNA is present at each. From FISH experiments on metaphases prepared from LB20 at both times of culture, there seems to be a mixed population of clones. One clone type shows a shortened chromosome 11q, with a site of integration in 11p, and the other shows full length human chromosome 11 with an integration site on 11q detected on the first culturing of the clone.

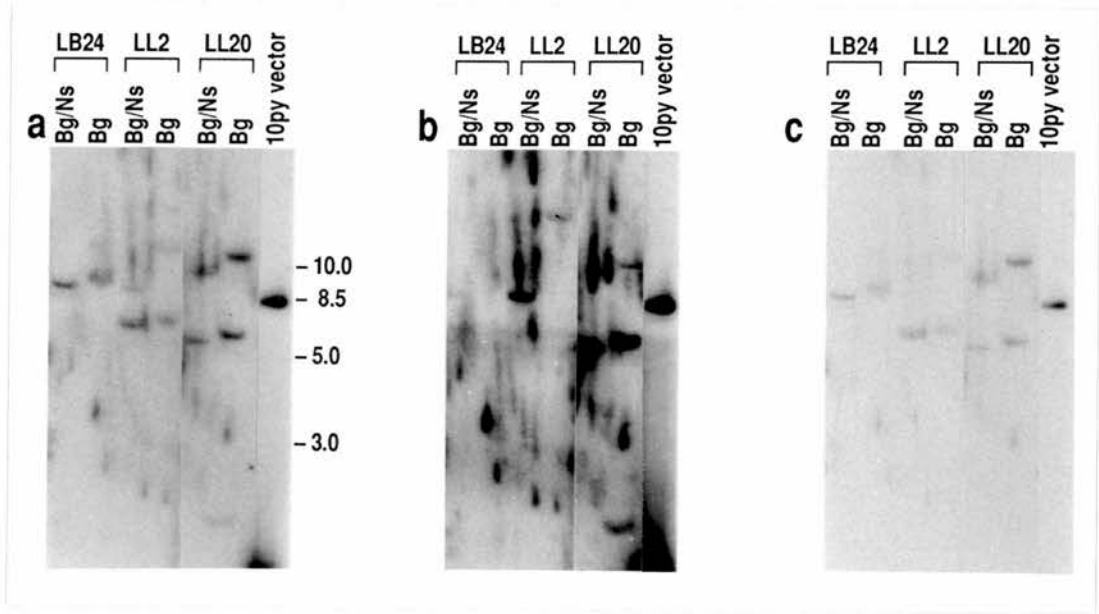
It is possible that the quantity of DNA from this clone on these blots is insufficient to enable an integration site present in 50% of the cells to be detected by the pBS probe, as we would have expected the integration sites to contain pBS sequences if recombination into the human chromosome has been directed by the repeat elements. It is unlikely that both these recombination events were random and that the vector integrated into the human chromosomes in both the cell types of this clone by chance, however if this were the case then the integration site will not necessarily contain any of the sequences in the construct apart from the *neoR* gene for which selection was applied.



Figure 5.9 Restriction maps of integration sites in clones LB5 and LL2 (not to scale)



**Figure 5.10** Restriction analysis of clones LB24, LL2, and LL20



A blot of BglIII and BglIII/NsiI digested DNA from clones LB24, LL2, and LL20 was prepared and hybridised exactly as described for Fig. 5.7. The results shown are (a) hybridisation with pTVLOXN- (probe X); (b) hybridisation with Bluescribe DNA (probe Y); and (c) hybridisation with *neoR/Leu2* fragment (probe Z).

### Clone LB24

Fig. 5.6 (a) shows that a single NsiI fragment is detected by probe X (pTVLOXN<sup>-</sup>) in DNA from clone LB24 which does not hybridise with probe Y (pBS), shown in Fig. 5.6 (b). This suggests that there is a single site of integration in this clone. Fig. 5.10 shows two BglII fragments hybridising to probes X and Z (*neoR/Leu2*) ((a) and (c)), neither of which are detected by probe Y (b). Surprisingly, a ~8kb BglII/NsiI band detected by probes X and Z is also detected by probe Y, albeit rather faintly (Fig. 5.10 (b)). Perhaps this integration site contains only a small fragment of pBS DNA such that its detection by hybridisation is very sensitive to the quantity and quality of genomic DNA in the sample. No clear conclusions can be drawn from this data.

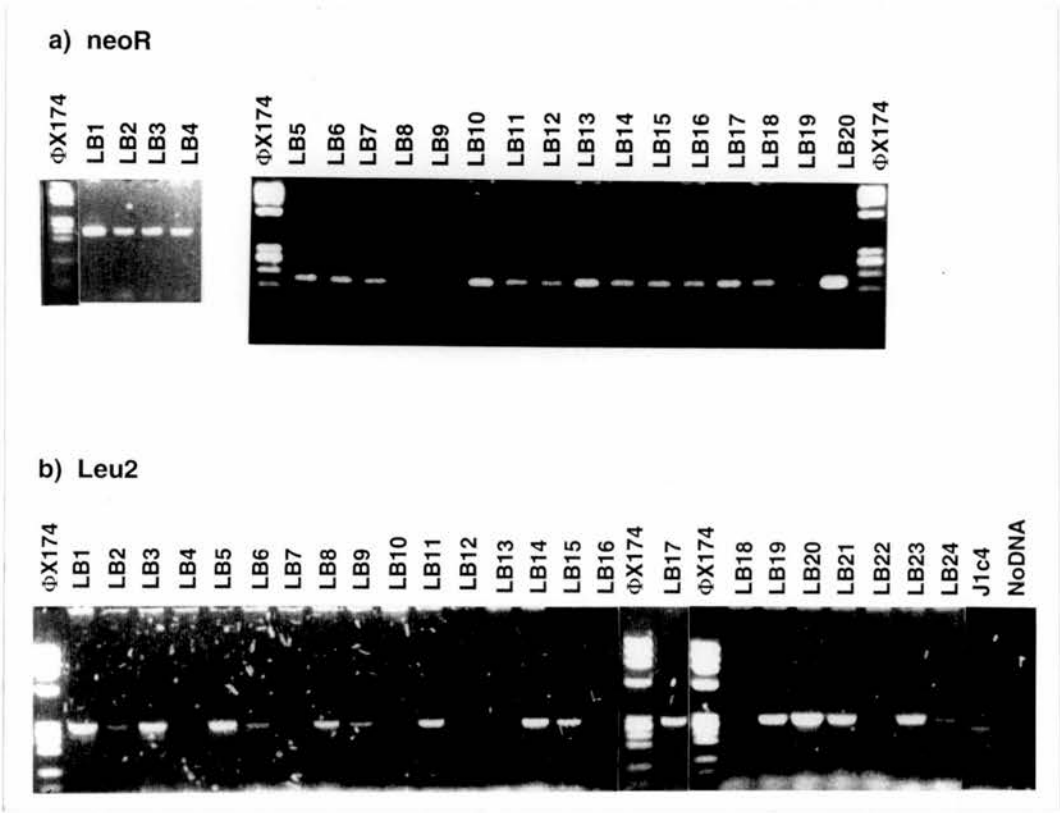
### Clone LL2

Fig. 5.6(a) shows that clone LL2 contains two sites of integration as two NsiI fragments, one of 16kb and one of ~7kb are detected by probe Z (*neoR/Leu2*). Only the upper band of ~16kb is detected by probe Y (pBS) (b), suggesting that this fragment encompasses an integration site containing a larger fragment of vector DNA. BglII/NsiI double digests give fragments of 8.5kb detected by probe Y and 7kb detected by probe Z, both of which are detected with probe X (pTVLOXN<sup>-</sup>), shown in Fig. 5.10 (a), (b), and (c). Probes X and Z also detect a 6.5kb BglII/NsiI band which does not hybridise with probe Y. This suggests that the 6.5kb NsiI fragment detected by probes X and Z does not contain a BglII site, and is likely to be a separate integration site. A single BglII fragment of ~19kb is detected by probe Y, fragments of 7kb and 11.7kb are detected by probe Z and all three of these fragments are detected by probe X. A likely map of the two integration sites in clone LL2 has been constructed from the restriction digest data and is shown in Fig. 5.9. The presence of an I-SceI site was not confirmed.

### Clone LL20

Probe Z (*neoR/Leu2*) detects three NsiI fragments in LL20 DNA; ~14kb, ~10.7kb and ~7.5kb, shown in Fig. 5.6(a), indicating that there are three sites of integration in this clone. Two of these sites contain pBS sequences, shown in Fig. 5.6 (b) to be within the ~14kb and ~7.5kb NsiI fragments. The 7.5kb fragment probably contains multiple copies of the pBS sequence, shown by the intensity of the bands detected by probe Y (pBS) (Fig. 5.6 (b)). 7.5kb is too small to encompass a tandem array of vector sequences so there has probably been a complicated recombination and integration event of the vector which has resulted in multiple copies of the pBS sequence being present at this integration site. The ~5.7kb BglII fragment and the ~5.6kb BglII/NsiI fragment shown in Fig 5.10 (a), (b), and (c) to be detected by all three probes probably originates from this same site as it again shows a high intensity

Figure 5.11 Detection of *neoR* and *Leu2* in pLL and pLB targeted clones by PCR



PCR was performed on ~100ng genomic DNA from clones LB1 to LB24 and clones LL2 to LL20. The primers and reaction conditions used are described in section 2.13.3. 10µl of a 20µl reaction was loaded onto a 2% nuseive gel and subjected to electrophoresis in 1xTBE at 90v for ~30min. The gel was stained with ethidium bromide and photographed. (a) shows the results of PCR with primers specific for the *neoR* gene with DNA from clones LB1 to LB20. (b) shows the results of PCR with DNA from clones LB1 to LB24 using primers specific for the *Leu2* gene.

of signal when hybridised with probe Y. One of the other integration sites contains two BglII fragments detected by probe Y, one of ~11.4kb and one of ~2kb. If recombination had occurred in the predicted fashion then only a single BglII fragment would be detected by probe Y per integration site. A definitive map could not be drawn up for these sites of integration without further restriction analysis.

## 5.2 Analysis of pLL- and pLB-targeted clones by Polymerase Chain Reaction (PCR)

Conditions were also developed to screen clones by PCR to detect those containing specific sequences introduced by the vector construct. It was reasoned that if recombination had been mediated by the repeat sequences then vector sequences lying between them should be present at the integration site and would be detectable by PCR.

Conditions were optimised for PCR with oligonucleotides from the 3' end of the *neoR* gene and the 3' end of the *Leu2* gene. The primer sequences and reaction conditions are given in section 2.13.3. The predicted product sizes were ~212bp for *neoR* and ~300bp for *Leu2*. Samples of the results of the reactions are shown in Fig. 5.11, and the results are summarised in Table 5.2.

From Table 5.2 it can be seen that PCR gave the required product with *neoR* primers in all the pLL and pLB clones. This was as expected as the clones carried resistance to G418 and therefore should have the *neoR* gene. Not all clones, however, gave bands with the *Leu2* primers.

Ideally, we wanted to identify clones containing Bluescribe sequences as these sequences lie adjacent to the 5' Line1 repeat fragment in pLL and pLB and it would therefore be a good way of identifying potential homologous recombination events. Identification of clones in this way would therefore be a rapid method of enriching for those in which integration of selectable markers has been mediated by the repeat sequences and hence those which have a tagged human chromosome. Unfortunately PCR using primers derived from Bluescribe sequence was particularly prone to contamination in the laboratory from air-borne plasmid DNA molecules, so that PCR consistently gave products in the negative control sample which was not supposed to contain any DNA. In addition, Southern blot analysis of these clones revealed that some of the DNA samples were contaminated with plasmid DNA. This would also affect the result of PCRs. Restriction enzyme digestion and Southern blot hybridisation was therefore adopted to screen for pBS at integration sites.



**Table 5.2 Summary of molecular analysis of isolated clones: results of PCR to detect sequences from neoR and Leu2 genes within targeted clones and results of BglII digest/hybridisations**

Clone	PCR analysis		Blot/ hybridisation data		FISH result
	neoR	Leu2	BglII fragment(s) in ~kb	pBS positive	
LB1	+	+	2.8, 5	+	human
LB2	+	+	10.2	+	n/d
LB3	+	+	9.3	+	n/d
LB4	+	-	4.8	-	n/d
LB5	+	+	16, 8	+	CHO
LB6	+	+	6.3	-	n/d
LB7	+	-	7.5	-	n/d
LB8	+	+	8.4	-	n/d
LB9	+	+	16	-	n/d
LB10	+	-	18	-	n/d
LB11	+	+	2.6, 6.4	+	n/d
LB12	+	-	?	?	n/d
LB13	+	-	8.0	-	n/d
LB14	+	+	4.7	-	n/d
LB15	+	+	14.4	+	human→CHO
LB16	+	-	4.5, 8.0	-	n/d
LB17	+	+	4.7, 7.2	-	n/d
LB18	+	-	11.4	-	n/d
LB19	+	+	4.7	-	n/d
LB20	+	+	(4), 5.9, 7.9, 9	+	human
LB21	+	+	4.7	-	n/d
LB22	+	-	?	?	n/d
LB23	+	+	6	+	n/d
LB24	+	+	9.0, 10.4	+	human

Table continues over the page

Table 5.2 contd.

Clone	PCR analysis		Blot/hybridisation data		FISH result
	neoR	Leu2	BglII fragment(s) in ~kb	pBS positive	
LL2	+	+	(4), 6.3, 11.7, 19	+	human
LL3	+	+	3.5	-	n/d
LL4	+	+	3.3	-	n/d
LL5	+	+	11	-	n/d
LL6	+	+	7.7	-	n/d
LL8	+	-	6.5	(+)	n/d
LL9	+	-	4.2, 5.3	-	n/d
LL10	+	-	?	?	n/d
LL12	+	+	12	-	n/d
LL14	+	+	12	-	n/d
LL15	+	+	4.6	+	n/d
LL19	+	+	4.5	+	n/d
LL20	+	+	4.2, 6, 12	+	CHO
LL22	-	-	3.8	-	n/d
J1c4	-	-	-	-	n/a
1.22	+	+	7.7	n/d	human

Key

( ) = band likely to have arisen from plasmid contamination

? = no clear result obtained

n/d = experiment not done

n/a = experiment not applicable

### 5.3 Summary of restriction enzyme analysis and PCR data for LB and LL clones

Molecular analysis of clones was carried out by restriction digestion followed by Southern blotting and hybridisation, and also by polymerase chain reaction (PCR). If recombination had occurred through the repeat elements of the targeting vector then it was expected that the intervening region from the plasmid construct would have inserted into the site of integration i.e. all sequences which lie between the two repeat elements in the linearised targeting construct (see Fig. 5.5) would be present in the targeted clones. This was tested as follows:

- a) all the vectors which were designed to insert the entire Bluescribe sequences into the integration site contain a BglII site in the 3' end of the *Leu2* gene, therefore DNA from clones in which the predicted insertion event has occurred should contain at least two BglII fragments which hybridise to a probe of vector DNA (pTVLOXN<sup>-</sup>);
- b) at least one of these BglII fragments should also hybridise to pBS sequences which are present in the vector;
- c) PCR using primers specific for the 3' end of the neoR gene and the 3' end of the *Leu2* gene should give a product of a designated size when performed on DNA from clones in which the predicted insertion event had occurred.

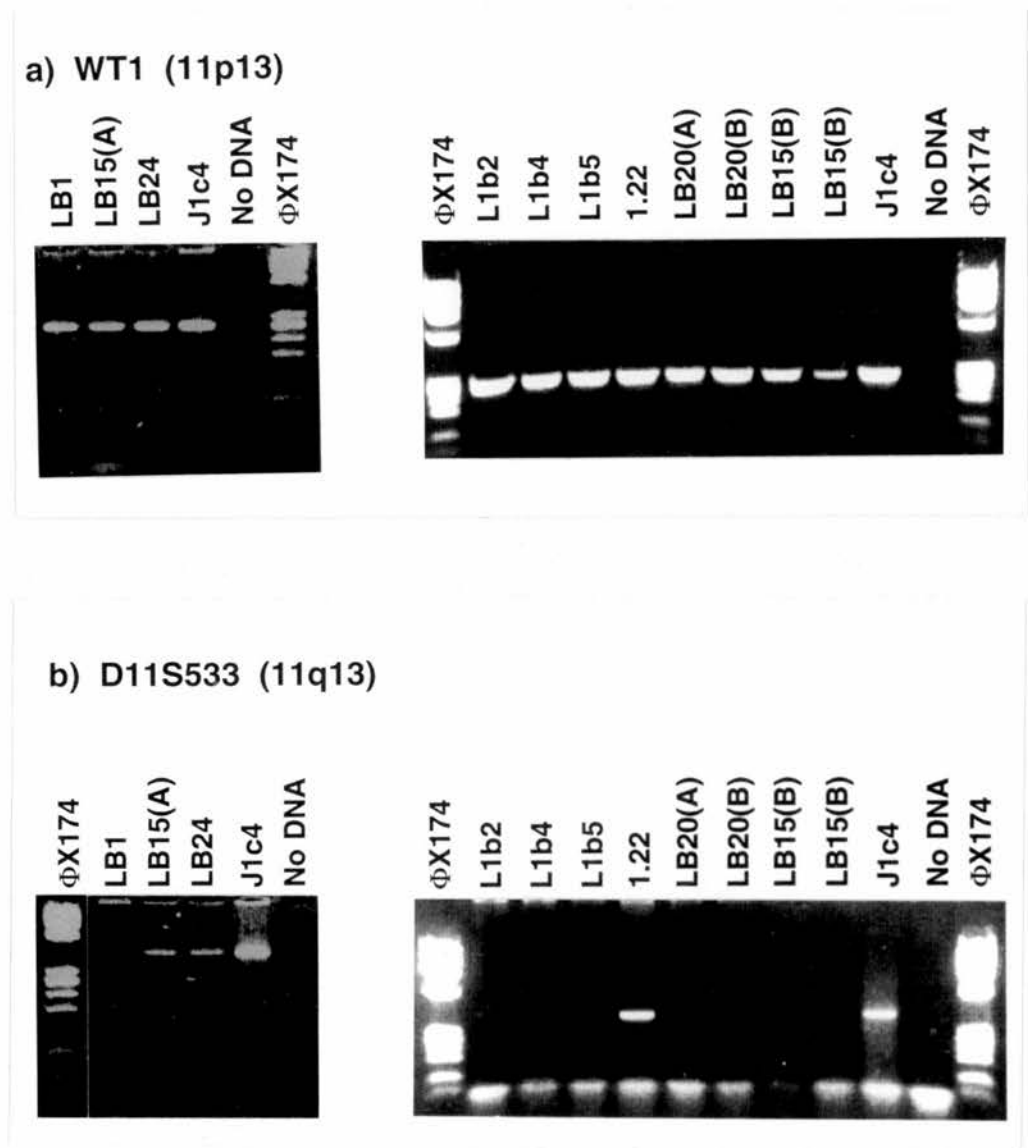
The results are summarised in Table 5.2. From the table it can be seen that:

- a) all clones contain at least one BglII fragment, but only 10 out of 35 clones for which data were obtained contain more than one BglII fragment,
- b) 12 out of 35 had a BglII fragment that contained pBS sequences,
- c) by PCR, all clones contain *neoR* but only 26/38 contain the *Leu2* gene
- d) all clones positive for pBS were also positive for *Leu2* indicating that no internal rearrangement of the vector had occurred in these clones, or a complicated rearrangement and a tandem or double integration event had occurred.

These data show that 7/35 clones fulfil the criteria for having the region intervening the repeat sequences introduced by the integration of pLL or pLB. In 4/6 of these clones analysed by FISH, the integration site was in the human chromosome. More analysis to localise the sites of integration for clones which were negative for the *Leu2* gene and pBS sequences for comparison would strengthen the argument for a correlation between the conservation of vector sequences and human integration sites.

Overall, these data show that a vector containing a Line1 fragment can mediate species-specific recombination and that these events can be enriched for by screening for clones containing sequences lying between the repeat elements within the vector. The molecular analysis does not clarify whether or not integration of L1-containing vectors into CHO DNA occurs by recombination with L1-Cg repeats or

Figure 5.12 Detection of the loci WT1 and D11S533 in selected clones by PCR



PCR was performed on ~100ng of genomic DNA from selected clones. The primers and reaction conditions used were as described in section 2.13.3. 10µl of a 20µl reaction was loaded onto a 2% nuseive gel and subjected to electrophoresis in 1xTBE at 90v for ~30min. The gel was stained with ethidium bromide and photographed. (a) shows the results of PCR with primers specific for the WT1 locus on human chromosome 11p13 with DNA from selected clones. (b) shows the results of PCR performed on DNA from the same clones with primers specific for the locus D11S533 located on the human chromosome 11 at 11q13.

whether it occurs through a random recombination mechanism. Future experiments for further investigation of the nature of these events are discussed in section 6.4.

5.4. Identification of fragments lost from the human chromosome in targeted clones

In order to identify which arm of the human chromosome 11 had been lost in those clones which showed a changed karyotype two PCR reactions were performed on samples of DNA from the clones. A set of oligonucleotide primers from the WT1 locus (a gift from K. Williamson) was used to detect sequences from 11p13 and a set of primers for the locus DS11533 (a gift from K. Evans) was used to detect sequences from 11q13.5. Oligonucleotide sequences and PCR conditions are given in section 2.13.3. A sample of the results is shown in Fig. 5.12. The data are summarised in Table 5.3.

Table 5.3 Summary of PCR results to identify p or q arm loss from the human chromosome 11 in selected clones

Clone	Targeting vector	Site of integration	WT1 PCR (11p13)	DS11533 PCR (11q13)	Cytologically-shortened chromosome11?
L1b2	L1b	CHO	+	+	no
L1b4	L1b	human	+	-	yes
L1b5	L1b	human	+	-	yes
1.22	pTVBlur1	human	+	+	no
LB1	pLB	human	+	-	yes
LB24	pLB	human	+	+	no
LB20	pLB	mixed	+	+(reduced)	50%
LB15(A)	pLB	human	+	+	no
LB15(B)	pLB	CHO	+	+(reduced)	50%
J1c4	none	none	+	+	no

From the table it can be seen that all clones which were observed to have a shortened human chromosome 11 by *in situ* analysis were negative for the q arm locus D11S533 by PCR. Clones containing a mixed population of cells, some with a

shortened chromosome 11 and some with a normal 11, showed a reduction in the amount of product obtained with PCR of the q arm locus relative to the amount obtained with PCR of the p arm locus (WT1). It can be concluded therefore that in all clones examined which had a shortened human chromosome 11, the q arm of the chromosome had been lost. This is discussed in more detail in Chapter 6.



## **CHAPTER 6**

## 6 Discussion and Conclusions

### 6.1 Integration of the vectors: analysis by FISH

#### 6.1.1 Sites of integration

The aim of this project was to test the ability of different human interspersed repeat repeats to mediate species-specific recombination within a somatic cell hybrid.

Vectors were constructed which contained the *neoR* selectable marker located between two cloned Alu repeats (Blur8 elements), two Line1 repeats or between an Alu and a Line 1 repeat. The constructs were transfected into the hybrid cell line J1c4 which is a Chinese hamster / human monochromosome hybrid containing a single human chromosome 11. Resulting clones were selected for G418 resistance and then analysed by FISH to localise the site of integration of the selectable marker.

Eighteen clones were analysed from a transfection experiment with a vector containing 2 Alu repeats in a head-to-head orientation (pTVBlur1) and 19 were analysed from an experiment with a vector in which 2 Alu repeats were in a head-to-tail configuration (pTVBlur2). FISH analysis showed that in only 1 out of the total of 37 clones produced from the Alu 'targeting' experiment had the vector integrated into human DNA. The ratio of human DNA to CHO DNA in J1c4 is ~1:28 so the observed frequency of 1 in 37 is within the range expected if integration of the vector was occurring entirely at random with a high degree of probability ( $p > 0.90$ ). Similar results were obtained when the same vectors were used to transfect two mouse fibroblast cell lines containing human chromosome 17 as its sole human component (GM10498 and GM10321). The ratio of human to rodent DNA in these hybrids is approximately the same as that for J1c4. Out of 40 clones screened by FISH (analysis performed by M. Stack) in only one was the vector observed to have inserted into a human chromosome. We can conclude therefore, that two Blur8 cloned Alu repeats elements in either tandem or opposite orientations do not promote human-specific recombination of a vector containing selectable markers within the two types of somatic cell hybrid cell lines tested. This agrees with the results of Wallenburg *et al.* (1987) as described in section 1.6.2.

In contrast, using a vector (pL1b) previously constructed by A. J. Brookes containing two Line1 repeat fragments to transfect J1c4, 3 out of 8 clones analysed by FISH showed integration of the vector into human chromosome 11. Vectors containing either two L1 repeat fragments (pLL) or one L1 fragment and the Blur8 repeat (pLB) were then built on a new design so that Bluescribe sequences would be introduced along with the selectable markers at the site of integration. Clones resistant to G418 created by transfection of these vectors into J1c4 cells were screened to

identify those containing Bluescribe sequences which lie between the two repeat elements in the vectors. FISH analysis showed that 1/2 and 3/5 of the selected clones, produced by pLL and pLB respectively, had human chromosome 11 integration sites. Another of the five pLB clones had an integration site in a human chromosome, which rearranged or was part of a sub-population of cells which was lost, after freezing and thawing of the culture (discussed below), giving a total of 5/7 human integrants for these experiments.

The original vector pL1b containing two Line1 repeat fragments gave a higher frequency of targeted integration events (3/8) than either of the redesigned vectors pLL and pLB, for which the frequency of 4/7 was obtained after pre-screening by Southern analysis. This could be because pL1b was a more conventional positive-negative selection vector which more readily promoted recombination through the repeat sequences. Vectors pLL and pLB were designed so that an extra 3kb of plasmid DNA would be introduced at the site of integration which may have reduced the frequency of homologous recombination events. There are insufficient data to conclude anything about the effect of having two Line1 fragments compared to only one upon the efficiency of vector integration.

These data suggest that a vector containing Line1 repeat fragment(s) integrates preferentially into a human chromosome 11 in the somatic cell hybrid J1c4 compared to a vector containing two Alu repeats which integrates at random into both human and Chinese hamster chromosomes. The significance of these frequencies is discussed in section 6.3.

#### *6.1.2 Chromosome shortening and stability*

In some metaphase preparations of the G418-resistant clones arising from these transfections it was observed that the human chromosome 11 was shortened so that one arm was almost completely absent. These clones were analysed by PCR using primers for a locus on each of the p and q arms to determine which had been lost. In all 9 clones tested, the WT1 locus at 11p13 was present. All 3 clones with a short human chromosome were negative for the locus DS11533 at 11q13. In those clones which contained a mixed population of cells with some containing a short chromosome 11 and others with a cytologically normal chromosome 11, the PCR product using primers from the 11q13 locus was correspondingly reduced. It has been previously documented that a region encompassing the loci HRAS, INS, TH, and H19 is selectively retained in deletion hybrids derived from J1 (Glaser *et al.* 1989). It has been proposed that this sub-telomeric region of human chromosome 11p contains a locus which complements a recessive lethal mutation of the host CHO-K1 cell line and

confers stability on the J1 somatic cell hybrid. This region has been mapped in the normal hamster diploid genome to parts of the short and long arms of chromosome 3 (Glaser *et al.* 1989). This would explain why human 11p is always retained but 11q may be lost in sub clones arising from these transfection experiments.

Previous observations by P. Byrd and D. Lawson (unpublished) of the karyotype of J1c4 in culture in our laboratory have shown that in most samples, ~5% of cells contain the human chromosome 11 derivative lacking 11q. Aliquots of the clone to be transfected with the vector constructs were assessed for heterogeneity and loss of 11q by scoring approximately 50 metaphases (work done by D. Lawson). Those samples transfected routinely contained <5% of cells with a shortened human chromosome. However after tagging of the clones, the proportion without 11q increased to ~30% in all the transfection experiments observed. This suggests that cells with the derived human chromosome are enriched for in some way during the transfection and selection procedure. Perhaps losses of a locus or loci present on 11q renders clones more easily transfected by lipofection by affecting the permeability of the cell nucleus for example, or alternatively makes them more resistant to G418, either through an increase in the effectiveness of the *neoR* gene product or some other property of the cell. An increased plating efficiency and/or early phase growth of 11q-derivatives might be an alternative explanation, but an overall increase in rate of cell division or viability is not consistent with the observation that the proportion of 11q-cells does not rise above ~5% under normal culture conditions. Clones in which the human chromosome 11 was shortened were observed in approximately equal proportions in all the transfection experiments, so there does not appear to be an enrichment for homologous recombination events within cells in which human chromosome 11q has been lost.

Differences in the morphology of the human chromosome or the location of the tag were observed after successive expansions of the clone LB15. This could be due to the primary clone being composed of different sub-populations, one of which survives stress or selection better than others, resulting in different proportions of the different sub-populations in successive sub-cultures of the clones. A mixed clone could have arisen during the initial isolation of the clone when it was first picked. There is some evidence for the existence of a mixed primary culture from the analysis of clone LB20, in which two different cell types present in approximately equal proportions were observed in the primary culture.

Alternatively, secondary populations could arise subsequent to the integration event. Recombination of the vector with a hybrid chromosome may induce chromosomal rearrangement or breakage. Strategies have been developed to enhance



and stabilise the breakage of chromosomes in mammalian cells by introducing a human telomere carried by the vector into the site of integration (Farr *et al.* 1992, Itzhaki *et al.* 1992). This process has been termed telomere-associated chromosome fragmentation (TACF) by Farr *et al.* (1992), and its application to chromosome analysis is discussed later.

Selection for an endogenous drug resistance gene has been shown previously to induce chromosomal rearrangements in hamster cells, which increase the level of expression of the gene by amplification of the locus, for example selection for the drug-resistance gene carbamoyl-P synthetase, aspartate transcarbamylase, dihydroorotase (CAD) (Smith *et al.* 1990). There is some evidence of chromosomal rearrangement in clone LB1 shown in Fig. 4.7. An unusual fusion event has occurred giving rise to a dicentric chromosome 11 in some cells. This may have occurred through recombination between integrated vector sequences on chromosomes in different cells after an endoreduplication event. Dicentric chromosomes are inherently unstable, breaking during cell division unless one of the centromeres is inactivated. Inactivity of one of the centromeres has been observed in stable, segregating dicentric chromosomes in mammalian cells (Merry *et al.* 1985). The dicentric chromosome observed in these Chinese hamster cells could be an intermediate in a bridge-breakage-fusion process first described by McClintock in 1942 and more recently observed in Syrian hamster cells (Smith *et al.* 1992). In Syrian hamster cells, fusion near the telomeres of the B9 chromosome either between sister chromatids or between homologous chromosomes was proposed by Smith *et al.* (1992) to lead to formation of a dicentric chromosome. Unequal breakage of the dicentric chromosome will result in both copies of a gene located close to the point of fusion segregating into one daughter cell, or segregation of the entire dicentric into one cell will result in duplication of all the genes on the chromosome. This was proposed as the mechanism resulting in amplification of the CAD gene. Chromosome fusion and breakage is also proposed as a mechanism for amplification of the endogenous DHFR gene in CHO cells under methotrexate selection (Windle *et al.* 1991, Windle and Wahl 1992). Kaufman *et al.* (1983) also proposed a mechanism of fusion and breakage to explain the amplification of an exogenous DHFR gene transfected into DHFR<sup>-</sup> CHO cells. It is possible such peri-telomeric recombination has occurred either between the sister chromatids of the tagged human chromosome in clone LB1 or between fortuitously aligned homologous chromosomes, induced by integration of the plasmid vector. This could be an intermediate in a pathway leading to amplification of the number of copies of the *neoR* gene. Resulting clones could then be selected for by their increased resistance to G418. This is thought to be a rare phenomenon in this study, however,

since a dicentric chromosome was only observed in a few cells of this one clone out of all the integration events analysed. From the molecular analysis of the sites of integration in a sample of the clones there is evidence for changes in the sites of integration in successive expansions in only one of these clones (LB15).

J1c4 is a hybrid cell line whose karyotype has been observed to be remarkably stable through prolonged periods of culture under normal conditions. It has been well characterised and has been used as part of a hybrid panel for the assignment of markers to chromosome 11 by many groups. Stable derivatives, selecting for or against cell surface markers on 11p have also been established (Glaser *et al.* 1989). It is therefore considered likely that variations in cell type within a tagged clone are due to a mixed primary cell population and not a consequence of genomic rearrangements after transfection. In a mixed population, clones which have lost human chromosome 11q may withstand G418-selection more effectively and therefore have an initial growth advantage over those which retain 11q, leading to different proportions of the sub-populations in successive expansions of the clone.

## **6.2 Integration of the vectors: molecular analysis**

From FISH analysis of the clones arising from these transfection experiments it is proposed that Line1 fragments in the plasmid constructs are causing species-specific recombination by promoting a recombination event with homologous sequences present in the human chromosome in the hybrid cell. We wanted to investigate the nature of this recombination event to find out if homologous recombination was actually occurring. It had been noticed that not all clones arising from L1-targeting experiments had integration sites in the human chromosome so we also wanted to investigate if these events were due to random recombination or recombination with CHO repeat sequences analogous to the human repeat.

More detailed restriction analysis performed on clones produced by the vectors pLL and pLB, for which the integration site had been localised did not produce any further evidence for or against homologous recombination. Most of the clones analysed seemed to contain multiple integration sites and this did not always correlate directly with the number of sites detected by FISH. Some of the sites of integration contain only a small region of vector DNA which is approaching the limit of detection possible using conventional fluorescent hybridisation of metaphase chromosomes, so it is possible that not all vector DNA present in the genome of J1c4 was detected. Double digests with an enzyme cutting internal to the vector and one cutting externally did not give patterns that could be readily analysed. Unequivocal analysis of recombination events is extremely difficult when a restriction map of the genomic



target is unknown. The hybridisation results were also complicated by the multiple integration sites and the presence of contaminating plasmid DNA in some samples of genomic DNA from the clones.

The region intervening the repeat sequences of the vector was, however, found to be intact in some of the clones and some contained a restriction enzyme fragment pattern predicted for a homologous recombination event. Enough data were obtained to enable possible restriction maps of some of the integration sites to be drawn, and to identify restriction fragments suitable for plasmid rescue. Circularisation and ligation of those NsiI fragments containing pBS sequences should allow them to be propagated after transformation into bacteria. The genomic sequences from one or both sides of the insertion point of the vector could be obtained from such plasmids. This would give us a clearer picture of the nature of the recombination event, as discussed later.

The I-SceI site was conserved in some of the clones and was shown to be functional. Digestion of the clones with I-SceI alone gave rise to unresolved fragments which is predicted for a unique cleavage event, and produced expected specific fragments when the clone DNA was also digested with BglII. This demonstrates that an I-SceI site introduced by these vectors into a mammalian hybrid cell genome is recognised and cleaved by the I-SceI restriction enzyme. This means that long range maps using I-SceI in combination with rare-cutting restriction enzymes such as NotI can be produced around the site of integration to facilitate the ordering of probes in that region. This is discussed further in section 6.5.

### **6.3 Line-1 elements and species-specific recombination**

With vectors containing L1 repeat fragments, species-specific recombination was observed by FISH in a high proportion of clones which were positive for pBS sequences (5/7). However, only 7/35 clones had a restriction digest hybridisation pattern compatible with a straight-forward homologous recombination event which would introduce pBS sequences, and of 6 of these putative targeted events, 4 had human integration sites. If the predicted Southern blot band pattern is taken as a criterion for homologous recombination then the overall frequency with which vectors pLL and pLB introduce selectable markers into human DNA can be estimated as  $4/6 \times 7/35 = \sim 2/15$  (13.3%). Compared to a frequency of 1/28 expected for an entirely random recombination event this is significantly different ( $\chi^2 = 4.46$  with one degree of freedom using the Yates' correction) giving a probability of  $p < 0.05$ . This frequency is probably an underestimate as it only includes those clones for which the Southern blot data were unequivocal and excludes other clones with a tagged human chromosome, for example LB24. Another factor affecting the estimated frequency

which has to be taken into consideration is that if, as postulated, chromosome 11 derivatives lacking 11q are pre-existing sub populations of the clone, the proportion of human DNA available for targeting in these clones is reduced to  $\sim 1/84$  as the q arm comprises  $\sim 2/3$  of chromosome 11 (11p  $\sim 50$ Mb and 11q  $\sim 100$ Mb). One clone (LB20) consisted of a mixed population in which 50% of cells contained a tagged full length 11, and the other 50% contained a tagged der(11). With an estimate that  $\sim 1/3$  of tagged, G418-selected clones have the der(11) then the overall chance for a vector to integrate into human DNA in a transfection experiment is  $\sim 1/36$ . Including all clones shown by FISH to have a site of integration on the human chromosome, and considering LB20 as 2 separate clones then the observed frequency of human integrations is 6/36 or  $\sim 17\%$ . The difference between this frequency and that predicted for a random recombination event which takes into account the total amount of human material available for targeting is highly significant ( $\chi^2 = 20.82$  calculated as before) with a probability of  $p < 0.01$ . Even compared with the conservative recombination frequency predicted for random events of 1/28, 6/36 is a highly significant difference ( $\chi^2 = 14.33$  calculated as before) with a probability of  $p < 0.01$ .

This demonstrates that even without selection for clones containing possible homologous recombination events, Line1 repeats can give a significant enrichment for human-specific recombination events. The data for the vector pL1b are however even more convincing, with 3/8 (37.5%) of clones showing human integrations without any pre-screening. This is an enrichment of approximately 10 fold over the expected frequency of a random event (3.57%). It is therefore proposed that variation between L1 elements within the human chromosome 11 and those in the vectors is small enough and variation between L1-H and L1-Cg repeats is great enough for Line repeats to mediate species-specific recombination in approximately 17% of transfected J1c4 clones.

It has previously been shown that homologous recombination events can be enriched for using positive-negative selection (PNS) vectors, indeed when this was tested in experiments 1, 2 and 3 there were 10-40 times fewer clones surviving on double selection of G418 plus gancyclovir compared to those surviving on G418 alone, thus enriching for those clones which contain an intact *neoR* gene but have lost a functional HSV-tk gene. However, in experiments 5 and 6 no significant difference in the numbers of clones obtained on either G418 alone or G418 plus gancyclovir was observed. It is thought that negative selection did not work in these experiments with the PNS vectors due to prolonged exonuclease digestion of the linearised vector constructs which probably removed promoter elements at the 5' end of the tk gene. If the experiment were to be repeated, Bal31 treatment would be omitted to see if the

number of homologous integrants could be increased, or alternatively, the HSV-tk gene would be omitted and an insertion-type vector would be used. This is discussed in section 6.5.

Homologous events have been screened for to a certain extent by selection of clones surviving G418 selection in which vector sequences intervening the repeat elements of the construct are present: 4/6 of clones selected in this way had stable human integration events, an enrichment of ~19-fold over random recombination (increased to 42-fold if multiple integrations and reduced target size are taken into account). If the experiment was to be repeated on a large scale to generate, for instance, panels of hybrids with human chromosomes tagged at many different locations, some secondary enrichment could be made by screening for vector pBS sequences. This could be done by Southern blotting, as demonstrated here, or alternatively by PCR using primers specific for Bluescribe DNA. Using the latter approach great care has to be taken to avoid contamination of samples of genomic DNA with exogenous plasmid DNA which would confuse PCR results.

In conclusion then, this thesis demonstrates the potential of vectors containing a fragment or fragments of the L1 human repeat element for the introduction of selectable markers into human chromosomes within somatic cell hybrids. Additional screening for homologous recombination events can give a high degree of enrichment for hybrid clones in which the human chromosome is tagged.

#### **6.4 Future Studies**

Although these experiments have shown that a human repeat element can direct species-specific recombination they have also raised several questions as to the nature of the event. For instance, is homologous recombination actually occurring, and if so, is it with a subset of L1 repeats in human chromosome 11 which show greater homology to the targeting sequence in the vector? The latter question could be answered in part by the fact that although the 1.8kb fragment at the end of the targeting vectors was cloned from the 3' end of the  $\beta$ -globin locus which is located on 11p15 there is no evidence from the *in situ* hybridisation analysis to date that this site is preferentially 'targeted' by the vector. It would also be interesting to determine if there is a subset of L1-Cg repeats in CHO DNA showing sufficient homology with sequences in the vector to promote recombination. Additionally, the differences between the frequencies and patterns of integration events mediated by the two vector constructs, one containing two L1 fragments and one containing one L1 fragment and an Alu element, could be investigated further by analysis of more clones arising from the targeting experiments and a more detailed examination of the recombination



events. It would also be valuable to ascertain if the multiple sites of integrations of the vector into the human chromosome seen in some clones were due to the specific transfection conditions used in the particular experiments, or were a direct consequence of multiple homologous target sites for vector recombination. Previous experiments to target loci present in multiple copies has not shown a corresponding increase in the number of integration events (Zeng *et al.* 1990). Similarly, gene targeting in ES cells has not, to my knowledge, been documented as altering both alleles of a gene in a single transfection experiment, so evidence to the contrary would be worth investigating.

Direct molecular analysis by restriction mapping and PCR have proved inadequate methods for answering these questions, largely due to the multiple possible target sites for recombination. Plasmid rescue of sites of integration from both human and Chinese hamster chromosomes and subsequent cloning of the original genomic target from DNA of untransfected J1c4, would allow investigation of the nature of the recombination by analysis of the precise sequences involved. Finer localisation of the integration sites of tagged clones by high resolution *in situ* hybridisation would resolve whether or not specific loci on the human chromosome were being preferentially targeted. It may also show if there was a bias for recombination into R or G bands; G bands being enriched for L1 repeats, but R-bands having more open, and potentially 'recombinogenic' chromatin.

The use of a PNS vector in these experiments does not appear to have enriched for those recombination events in which the negative selectable marker has been lost in some experiments. The vulnerability of the negative selection gene to loss of function through mutation or degradation has been previously observed (Bernet-Grandaud *et al.* 1992). Future experiments could alternatively be undertaken with an insertion-type vector which does not contain an HSV-tk gene for negative selection, but instead has contiguous L1 fragments through which the vector is linearised. Any enrichment which may have occurred through negative selection may be compensated for by the higher frequency of homologous recombination observed with insertion-type vectors (Dickinson *et al.* 1993).

With J1c4, it has been shown that mixed populations of clones can arise. This is thought to be due to a non-clonal original population and not due to instability and rearrangement of the clones subsequent to integration. In future experiments, primary G418-resistant clones would be plated at very low concentrations in order to isolate colonies arising from a single cell. The individual secondary clones could then be screened first by PCR and then by FISH to identify those containing the required integration event. Clones should also be propagated for a prolonged period in culture

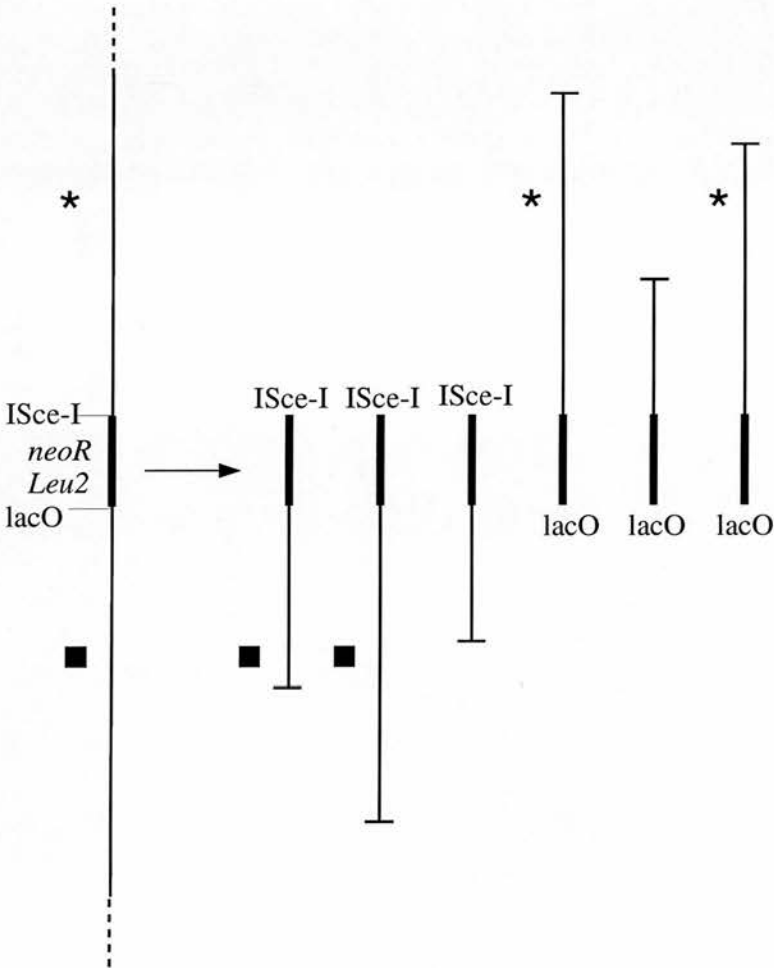
to check their stability. Purified and stable clones will then be available for subsequent applications.

### **6.5 Manipulation of hybrid clones using novel features introduced by the vector**

An I-SceI site introduced by these vectors has been shown to be recognised and cleaved by I-SceI restriction enzyme within samples of genomic DNA isolated from selected clones (see section 5.3.1, Fig. 5.8). It is proposed that the integration sites of these vectors will provide points for unique cleavage of a human chromosome. This has been previously demonstrated by Kurdi-Haidar *et al.* (1993) after a provirus was used to introduce I-SceI sites at random into the genome of mouse NIH/3T3 cells. I-SceI was used in double digests with rare cutting enzymes to generate a physical map spanning ~1Mb around a particular integrated provirus, this technique was termed provirus-anchored long-range (PAL) mapping. In comparison, repeat-mediated recombination will introduce unique cleavage sites into a specific human chromosome contained within a somatic cell hybrid. As well as allowing analysis of a specific chromosome, this technique allows the site of integration, and therefore the region to be mapped, to be first localised by FISH due to the length of vector sequence incorporated at the site of integration. A clone containing a unique cleavage site within a region of interest could be selected from a panel of hybrids in which the point of integration has been localised by FISH. Long range mapping of a region relative to a unique cleavage site will simplify the ordering of probes. Development of the Achilles' Heel method will enable a second site for unique cleavage of genomic DNA to be created at the lacO sequence introduced by the integrated vector. This has also been demonstrated by Kurdi-Haidar *et al.* (1993). Two different sites for unique cleavage, one at each end of the vector sequences will allow anchored mapping in both directions, as shown in Fig. 6.1.

DNA from an integration site in a region of interest could be rescued into *S. cerevisiae* using the selectable marker *Leu2* introduced by the vector. A specific fragment in the range of a megabase in size containing *Leu2* could be enriched using size selection or by affinity purification utilising the lacO system described in section 1.7.3. The selected genomic DNA can then be cloned into a YAC vector and transformed into yeast. Those clones surviving selection on leucine-deficient medium should contain the DNA of interest. Work in progress in the group has shown that fragments of genomic DNA containing the *Leu2* gene present in these vectors can be rescued from mammalian cells into yeast (J. Maule personal communication). Having a selected region of DNA in a YAC means that it can be analysed in detail physically, through restriction mapping and sub cloning. It can be also analysed functionally by

Figure 6.1 Anchored long range mapping with unique restriction sites



Digestion of genomic DNA is first with *ISce-I* or by the Achilles' Heel mechanism, at a unique site. Partial digests with a second enzyme, for instance the rare cutter *NotI*, will give rise to large overlapping restriction fragments. The fragments can be resolved by PFGE and ordered using vector sequences as a probe. Anonymous probes (\*) and (■) from around the region can then be assigned to the map.



combining existing techniques for the precise alteration of DNA cloned in yeast through homologous recombination, and techniques for introducing YACs back into mammalian cells where the exogenous DNA can be stably maintained and expressed, using the *neoR* selectable marker already incorporated (as discussed in section 1.7.4).

## **6.6 Applications for clones containing an exogenous selectable marker**

Using L1 repeats is an improvement on random strategies for introducing markers into human chromosomes in somatic cell hybrids. Methods using retroviral vectors (Lugo *et al.* 1987) or plasmid constructs (Koi *et al.* 1989) require very large numbers of clones to be screened to isolate those containing tagged human chromosomes as they do not promote species-specific recombination within a somatic cell hybrid.

Selectable markers in human chromosomes are essential for any strategy in which the identification, enrichment and often the stable maintenance of a particular chromosomal region is required. Used in conjunction with techniques for chromosome fragmentation and cell fusion such as CMGT and IFGT, cell lines with a reduced portion of the human chromosome can be produced from clones tagged by integration of the vector. Imposing G418 selection will ensure that the region of the human chromosome encompassing the original site of integration of the vector will be present in the secondary hybrid. Such 'reduced' hybrids can be characterised by Alu-PCR and FISH to become part of a hybrid mapping panel. They can also be used as resources for cosmid cloning and for the creation and mapping of novel probes from around the region of interest. Reduced hybrids have played an important part of positional cloning strategies which have led to the localisation of many novel human genes in the last 15 years, and this novel method for introducing selectable markers into human chromosomes should facilitate the creation of those containing selected and limited regions of the genome.

Repeat-mediated recombination could be combined with a technique to introduce specific chromosome breakage and shortening by including L1 repeats in a vector containing a human telomere. Farr *et al.* (1992) and Itzaki *et al.* (1992) have previously used telomere-containing constructs to cause chromosome breakage at specific loci by a targeted recombination event. A Line1 repeat could be used to introduce a telomere at several different loci thus producing several hybrid clones with the human chromosome shortened by different amounts in one transfection experiment. This strategy could be applied to any chromosome isolated within a somatic cell hybrid and is not limited to those with a negative selectable marker to enrich for human integration events. For example, the random integration approach of TACF described

by Farr *et al.* (1992) was successful in obtaining clones containing a truncated human X-chromosome because they could select for clones which had lost the HPRT gene on Xq. As well as a useful mapping resource, chromosomes with engineered breakages could provide insight into chromosome structure and stability, in particular in the analysis of human centromere and telomere requirements.

Full-length chromosomes or fragments induced by irradiation of the original clones containing a selectable marker can be used in microcell fusion experiments, as discussed in section 1.1. Microcells generated from the clones tagged by an integrated vector construct, could be fused with any cellular assay system, imposing G418 selection to enrich for those cells which had taken up the tagged exogenous DNA. For instance, cells from a patient with Ataxia telangiectasia can be used to identify those chromosomal fragments containing genes which can complement the UV-sensitivity of the cell line, as described by Lambert *et al.* (1991). Alternatively, chromosomal fragments could be assayed for genes encoding metabolic enzymes which affect the premature senescence shown by cultured fibroblasts from patients with Werner syndrome (Thweatt and Goldstein 1993). In other cellular systems they can be assayed for genes encoding certain cell surface antigens by the use of specific antisera, or they can be screened for the presence of tumour suppressor genes by their ability to change the growth patterns or tumorigenicity of tumour cell lines.

In conclusion, this method for somatic cell manipulation has several advantages over previously described techniques, and has the potential to make a valuable contribution to positional cloning and functional analysis of the human genome.

## References

- Anderson M. (1991) The multi-national coordinated *Arabidopsis thaliana* genome research project. *GNOME News* **8**, 26-27.
- Anderson R. A. and Eliason S. L. (1986) Recombination of homologous DNA fragments transfected into mammalian cells occurs predominantly by terminal pairing. *Mol. Cell. Biol.* **6**, 3246-3252.
- Anderson R. A., Kato S., and Camerini-Otero R. D. (1984) A pattern of partially homologous recombination in mouse L cells. *Proc. Natl. Acad. Sci. USA* **81**, 206-210.
- Andreadis A., Hsu Y. P., Hermodson M., Kohlhaw G., and Schimmel P. (1984) Yeast LEU2. Repression of mRNA levels by leucine and primary structure of the gene product. *J. Biol. Chem.* **259**, 8059-8062.
- Aruffo A. and Seed B. (1987a) Molecular cloning of a CD28 cDNA by a high-efficiency COS cell expression system. *Proc. Natl. Acad. Sci. USA* **84**, 8573-8577.
- Aruffo A. and Seed B. (1987b) Molecular cloning of two CD7 (T-cell leukemia antigen) cDNAs by a COS cell expression system. *Embo. J.* **6**, 3313-3316.
- Arveiler B. and Porteous D. (1992) Distribution of Alu and L1 repeats in human YAC recombinants. *Mammalian Genome* **3**, 661-668.
- Bains W. (1986) The multiple origins of human Alu sequences. *J. Mol. Evol.* **23**, 189-199.
- Baker S. J., Markowitz S., Fearon E. R., Willson J. K. V., and Vogelstein B. (1990) Suppression of human colorectal carcinoma growth by wild-type p53. *Science*. **249**, 912-915.
- Barski G., Sorieul S., and Cornefert F. (1960) Production dans des cultures in vitro de deux souches cellulaires en association, de cellules de caractere "hybride". *C. R. Acad. Sci. (Paris)* **251**, 1825-1827.
- Barski G., Sorieul S., and Cornefert F. (1961) 'Hybrid' type cells in combined cultures of two different mammalian cell strains. *J. Nat. Cancer Inst.* **26**, 1269-1291.
- Batzer M. A. and Deininger P. L. (1991) A human-specific subfamily of Alu sequences. *Genomics*. **9**, 481-487.
- Batzer M. A., Kilroy G. E., Richard P. E., Shaikh T. H., Desselle T. D., Hoppens C. L., and Deininger P. L. (1990) Structure and variability of recently inserted Alu family members. *Nucleic Acids Res.* **18**, 6793-6798.
- Benham F., Hart K., Crolla J., Bobrow M., Francavilla M., and Goodfellow P. N. (1989) A method for generating hybrids containing nonselected fragments of human chromosomes. *Genomics*. **4**, 509-517.
- Bernet-Grandaud A., Ouazana R., Morle F., and Godet J. (1992) A method improving the efficiency of the positive-negative selection used to isolate homologous recombinants. *Nucleic Acids Res.* **20**, 6417-6418.

- Bickmore W. A. and Sumner A. T. (1989) Mammalian chromosome banding--an expression of genome organization. *Trends. Genet.* **5**, 144-148.
- Bishop J. M. (1987) The molecular genetics of cancer. *Science.* **235**, 305-311.
- Bollag R. J., Waldman A. S., and Liskay R. M. (1989) Homologous recombination in mammalian cells. *Ann.Rev.Genet.* **23**, 199-225.
- Brenner D. A., Smigocki A. C., and Camerini-Otero D. (1985) Effects of insertions, deletions, and double-stranded breaks on homologous recombination in mouse L cells. *Mol. Cell. Biol.* **5**, 684-691.
- Brenner S., Elgar G., Sandford R., Macrae A., Venkatesh B., and Aparicio S. (1993) Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome. *Nature.* **366**, 265-268.
- Britten R. J., Baron W. F., Stout D. B., and Davidson E. H. (1988) Sources and evolution of human Alu repeated sequences. *Proc. Natl. Acad. Sci. USA* **85**, 4770-4774.
- Brookes A. J., Stevenson B. J., Porteous D. J., and Dorin J. R. (1993) A series of vectors that simplify mammalian gene targeting. *Transgenic. Res.* **2**, 238-244.
- Buerstedde J.-M. and Takeda S. (1991) Increased ratio of targeted to random integration after transfection of chicken B cell lines. *Cell* **67**, 179-188.
- Burke D. T., Carle G. F., and Olson M. V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science.* **236**, 806-812.
- Burton F. H., Loeb D. D., Voliva C. F., Martin S. L., Edgell M. H., and Hutchison C. A. 3d. (1986) Conservation throughout mammalia and extensive protein-encoding capacity of the highly repeated DNA long interspersed sequence one. *J. Mol. Biol.* **187**, 291-304.
- Call K. M., Glaser T., Ito C. Y., Buckler A. J., Pelletier J., Haber D. A., Rose E. A., Kral A., Yeger H., Lewis W. H., Jones C., and Housman D. E. (1990) Isolation and characterisation of a zinc finger polypeptide gene at human chromosome 11 Wilm's tumor locus. *Cell* **60**, 509-520.
- Capecchi M. R. (1989a) Altering the genome by homologous recombination. *Science.* **244**, 1288-1292. (review)
- Capecchi M. R. (1989b) The new mouse genetics: altering the genome by gene targeting. *Trends in Genetics* **5**, 70-76.
- Chang D.-Y. and Maraia R. J. (1993) A cellular protein binds B1 and Alu small cytoplasmic RNAs in vitro. *J. Biol. Chem.* **268**, 6423-6428.
- Chen T. L. and Manuelidis L. (1989) SINEs and LINEs cluster in distinct DNA fragments of Giemsa band size. *Chromosoma* **98**, 309-316.
- Chen Y., Chen P.-L., Arnaix N., Goodrich D., and Lee W.-H. (1991) Expression of wild-type p53 in human A673 cells suppresses tumorigenicity but not growth rate. *Oncogene* **6**, 1799-1805.



- Cobianchi F. and Wilson S. H. (1987) Enzymes for modifying and labelling DNA and RNA. *Methods Enzymol.* **152**, 94-110.
- Colleaux L., Rougeulle C., Avner P., and Dujon B. (1993) Rapid physical mapping of YAC inserts by random integration of I-SceI sites. *Nucleic Acids Res.* **2**, 265-271.
- Cook P. R. (1984) A general method for preparing intact nuclear DNA. *Embo. J.* **3**, 1837-1842.
- Cooper C. S., Park M., Blair D. G., Tainsky M. A., Huebner K., Croce C. M., and van de Woude G. F. (1984) Molecular cloning of a new transforming gene from a chemically transformed human cell line. *Nature.* **311**, 29-33.
- Cooper G., Okenquist S., and Silverman L. (1980) Transforming activity of DNA of chemically transformed and normal cells. *Nature.* **284**, 418-421.
- Coulson A., Waterson R., Kiff J., Sulston J., and Korhara Y. (1988) Genome linking with yeast artificial chromosomes. *Nature.* **355**, 184-186.
- Cox D. R. and Myers R. M. (1992) Bridging the gaps. *Curr. Biol.* **2**, 338-339.
- Cox D. R., Burmeister M., Price E. R., Kim S., and Myers R. M. (1990) Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science.* **250**, 245-250.
- Craig J. M. and Bickmore W. A. (1993) Chromosome bands - flavours to savour. *BioEssays* **15**, 349-354.
- Daniels G. R. and Deininger P. L. (1985a) Integration site preferences of the Alu family and similar repetitive DNA sequences. *Nucleic Acids Res.* **13**, 8939-8954.
- Daniels G. R. and Deininger P. L. (1985b) Repeat sequence families derived from mammalian tRNA genes. *Nature.* **317**, 819-822.
- Davidson R. L. and Ephrussi B. (1965) A selective system for the isolation of hybrids between L cells and normal cells. *Nature.* **205**, 1170.
- Deininger P. L. (1989) SINES: short interspersed repeated DNA elements in higher eukaryotes, in *Mobile DNA, 1st ed. Vol. 1* (Berg D. E. and Howe M. M., eds), pp. 619-636. Am. Soc. Microbiol., Washington DC, USA.
- Deininger P. L. and Schmid C. W. (1976) An electron microscope study of the DNA sequence organisation of the human genome. *J. Mol. Biol.* **106**, 773-790.
- Deininger P. L., Jolly D. J., Rubin C. M., Friedmann T., and Schmid C. W. (1981) Base sequence studies of 300 nucleotide renatured repeated human DNA clones. *J. Mol. Biol.* **151**, 17-33.
- Deininger P. L., Batzer M. A., Hutchinson C. A., and Edgell M. H. (1992) Master genes in mammalian repetitive DNA amplification. *Trends in Genetics* **8**, 307-311.
- Deng C. and Capecchi M. R. (1992) Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Mol. Cell. Biol.* **12**, 3365-3371.



- Deng C., Thomas K. R., and Capecchi M. R. (1993) Location of crossovers during gene targeting with insertion and replacement vectors. *Mol. Cell. Biol.* **13**, 2134-2140.
- Dickinson P., Kimber W. L., Kilanowski F. M., Stevenson B. J., Porteous D. J., and Dorin J. R. (1993) High frequency gene targeting using insertional vectors. *Hum. Mol. Genet* **2**, 1299-1302.
- Doetschmann T., Gregg R. G., Maeda N., Hooper M. L., Melton D. W., Thompson S., and Smithies O. (1987) Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature*. **330**, 576-578.
- Dombroski B. A., Mathias S. L., Nanthakumar E., Scott A. F., and Kazazian H. H. (1991) Isolation of an active transposable element. *Science*. **254**, 1805-1807.
- Dombroski B. A., Scott A. F., and Kazazian H. H. (1993) Two additional potential retrotransposons isolated from a human L1 subfamily that contains an active retrotransposable element. *Proc. Natl. Acad. Sci. USA* **90**, 6513-6517.
- Don R. H., Cox P. T., Wainwright B. J., Baker K., and Mattick J. S. (1991) 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* **19**, 4008.
- Dorin J. R., Inglis J. D., and Porteous D. J. (1989) Selection for precise chromosomal targeting of dominant marker by homologous recombination. *Science*. **243**, 1357-1360.
- Dorin J. R., Dickinson P., Emslie E., Clarke A. R., Dobbie L., Hooper M. L., Halford S., Wainwright B. J., and Porteous D. J. (1992) Successful targeting of mouse cystic fibrosis transmembrane conductance regulator gene in embryonal stem cells. *Transgenic. Res.* **1**, 101-105.
- Dowdy S. F., Scanlon D. J., Fasching C. L., Casey G., and Stanbridge E. J. (1990) Irradiation microcell-mediated chromosome transfer (XMMCT): the generation of specific chromosomal arm deletions. *Genes Chrom. Cancer* **2**, 318-327.
- Dowdy S. F., Fasching C. L., Araujo D., Lat K.-M., Livanos E., Weissman B. E., and Stanbridge E. J. (1991) Suppression of Tumorigenicity in Wilms Tumor by the p15.5-p14 Region of Chromosome 11. *Science*. **254**, 293-295.
- El-Deity W. S., Tokino T., Velculescu V. E., Levy D. B., Parsons R., Trent J. M., Lin D., Mercer W. E., Kinzler K. W., and Vogelstein B. (1993) WAF1, a potential mediator of p53 suppression. *Cell* **75**, 817-825.
- Ephrussi B. (1972) *Hybridization of Somatic Cells, 1st ed.* Princeton University Press, New Jersey.
- Ephrussi B. and Weiss M. C. (1965) Interspecific hybridization of somatic cells. *Proc. Natl. Acad. Sci. USA* **53**, 1040-1043.
- Farr C. J., Stevanovic M., Thomson E. J., Goodfellow P. N., and Cooke H. J. (1992) Telomere-associated chromosome fragmentation: applications in genome manipulation and analysis. *Nat. Genet.* **2**, 275-282.

- Fearon E. R. and Vogelstein B. (1990) A genetic model for colorectal tumorigenesis. *Cell* **61**, 759-767.
- Feinberg A. P. and Vogelstein B. (1983) A technique for radiolabelling DNA restriction fragments to a high specific activity. *Anal. Biochem.* **132**, 6-13.
- Feinberg A. P. and Vogelstein B. (1984) Addendum: A technique for radiolabelling DNA restriction fragments to a high specific activity. *Anal. Biochem.* **137**, 266-267.
- Felgner P. L., Gadek T. R., Holm M., Roman R., Chan H. W., Wenz M., Northrop J. P., Ringold G. M., and Danielsen M. (1987) Lipofection: a highly efficient, lipid-mediated DNA transfection procedure. *Proc. Natl. Acad. Sci. USA* **84**, 7413-7417.
- Finlay C. A., Hinds P. W., and Levine A. J. (1989) The p53 proto-oncogene can act as a suppressor of transformation. *Cell* **57**, 1083-1093.
- Folger K. R., Wong E. A., Wahl G., and Capecchi M. R. (1982) Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid molecules. *Mol. Cell. Biol.* **2**, 1372-1387.
- Fournier R. E. K. and Ruddle F. H. (1977) Microcell-mediated transfer of murine chromosomes into mouse, Chinese hamster, and human somatic cells. *Proc. Natl. Acad. Sci. USA* **74**, 319-323.
- Garrison P. N., Koob M., and Barnes L. D. (1992) Physical mapping of the *Saccharomyces cerevisiae* Ap4A phosphorylase I-encoding gene by the Achilles' cleavage method. *Gene* **111**, 131-134.
- Gerhard D. S., Kawasaki E. S., Carter Bancroft F., and Szabo P. (1981) Localization of a unique gene by direct hybridisation in situ. *Proc. Natl. Acad. Sci. USA* **78**, 3755-3759.
- Gerhard D. S., Jones C., Morse H. G., Handelin B., Weeks V., and Housman D. (1987) Analysis of human chromosome 11 by somatic cell genetics: reexamination of derivatives of human-hamster cell line J1. *Somat. Cell Mol. Genet.* **13**, 293-304.
- Gessler M., Poustka A., Cavanee W., Neve R. L., Orkin S. H., and Bruns G. A. P. (1990) Homozygous deletion in Wilm's tumors of a zinc-finger gene identified by chromosome jumping. *Nature*. **343**, 774-778.
- Glaser T., Housman D., Lewis W. H., Gerhard D., and Jones C. (1989) A fine-structure deletion map of human chromosome 11p: analysis of J1 series hybrids. *Somat. Cell Mol. Genet.* **15**, 477-501.
- Glaser T., Rose E., Morse H., Housman D., and Jones C. (1990) A panel of irradiation-reduced hybrids selectively retaining human chromosome 11p13: their structure and use to purify the WAGR gene complex. *Genomics*. **6**, 48-64.
- Gnirke A. and Huxley C. (1991) Transfer of the human HPRT and GART genes from yeast to mammalian cells by microinjection of YAC DNA. *Somat. Cell Mol. Genet.* **17**, 573-580.

- Gnirke A., Barnes T. S., Patterson D., Schild D., Featherstone T., and Olson M. V. (1991) Cloning and in vivo expression of the human GART gene using yeast artificial chromosomes. *Embo. J.* **10**, 1629-1634.
- Goold R. D., Dadgar J., diSibio G., Dugaiczky A., Lang D. B., Magrane G., Smith K. A., Xu H., Master S. B., Cox D. R., and Myers R. M. 1993, Development of 741 STS markers from human chromosome 4, Abstract, Cold Spring Harbor Genome Mapping and Sequencing.
- Goss S. J. (1978) Gene mapping by cell fusion, in *International review of cytology, supplement 8*, pp. 127-169. Academic Press Inc., .
- Goss S. J. and Harris H. (1975) New methods for mapping genes in human chromosomes. *Nature.* **255**, 680-684.
- Gossen J. A., de Leeuw W. J. F., Molijn A. C., and Vijg J. (1993) Plasmid rescue from transgenic mouse DNA using LacI repressor protein conjugated to magnetic beads. *BioTech* **14**, 624-629.
- Goyette M. C., Cho K., Fasching C. L., Levy D. B., Kinzler K. W., Paraskeva C., Vogelstein B., and Stanbridge E. J. (1992) Progression of colorectal cancer is associated with multiple tumor suppressor gene defects but inhibition of tumorigenicity is accomplished by correction of any single defect via chromosome transfer. *Mol. Cell. Biol.* **12**, 1387-1395.
- Grant S. G. N., Jessee J., Bloom F. R., and Hanahan D. (1990) Differential plasmid rescue from transgenic mouse DNAs into *Eschericia coli* methylation-restriction mutants. *Proc. Natl. Acad. Sci. USA* **87**, 4645-4649.
- Grimaldi G., Skowronski J., and Singer M. F. (1984) Defining the beginning and end of KpnI family segments. *Embo. J.* **3**, 1753-1759.
- Grimes E., Koob M., and Szybalski W. (1990) Achilles' heel cleavage: creation of rare restriction sites in lambda phage genomes and evaluation of additional operators, repressors and restriction/modification systems. *Gene.* **90**, 1-7.
- Gusella J. F., Keys, C., Varsanyi-Breiner A., Kao F. T., Jones C., Puck T. T., and Housman D. (1980) Isolation and localization of DNA segments from specific human chromosomes. *Proc. Natl. Acad. Sci. USA* **77**, 2829-2833.
- Hanahan D. (1983) Studies on transformation of *Eschericia coli* with plasmids. *J. Mol. Biol.* **166**, 557-580.
- Harris H. and Watkins J. F. (1965) Hybrid cells derived from mouse and man: artificial heterokaryons of mammalian cells from different species. *Nature.* **205**, 640-646.
- Harris H., Miller O. J., Klein G., Worst P., and Tachibana T. (1969) Suppression of malignancy by cell fusion. *Nature.* **223**, 363.
- Hastings P. J. (1988) Recombination in the eukaryotic nucleus. *BioEssays* **9**.
- Hasty P., Rivera-Perez J., and Bradley A. (1991a) The length of homology required for gene targeting in embryonic stem cells. *Mol. Cell. Biol.* **11**, 5586-5591.



- Hasty P., Rivera-Perez J., Chang C., and Bradley A. (1991b) Target frequency and integration patterns for insertion and replacement vectors in embryonic stem cells. *Mol. Cell. Biol.* **11**, 4509-4517.
- Hasty P., Rivera-Perez J., and Bradley A. (1992) The role and fate of DNA ends for homologous recombination in embryonic stem cells. *Mol. Cell. Biol.* **12**, 2464-2474.
- Haynes S. R., Toomey T. P., Leinward L., and Jelinek W. R. (1981) The Chinese hamster Alu-equivalent sequence: a conserved, highly repetitious, interspersed deoxyribonucleic acid sequence in mammals has a structure suggestive of a transposable element. *Mol. Cell. Biol.* **1**, 573-583.
- Heng H. H. Q., Squire J., and Tsui L.-C. (1992) High resolution mapping of mammalian genes by *in situ* hybridisation to free chromatin. *Proc. Natl. Acad. Sci. USA* **89**, 9509-9513.
- Houk C. M., Rinehart F. P., and Schmid C. W. (1979) A ubiquitous family of repeated DNA sequences in the human genome. *J. Mol. Biol.* **132**, 289-306.
- Hu M. C. and Davidson N. (1987) The inducible lac operator-repressor system is functional in mammalian cells. *Cell* **48**, 555-566.
- Hu X. and Worton R. G. (1992) Partial gene duplication as a cause of human disease. *Human Mutation* **1**, 3-12.
- Human Gene Mapping 8 (1985) *Cytogenetics and Cell Genetics*, Vol. 40, 1-4. Karger, Basel, Switzerland.
- Hutchinson III C. A., Hardies S. C., Loeb D. D., Shehee W. R., and Edgell M. H. (1982) LINEs and related retroposons: long interspersed repeated sequences in the eukaryotic genome, in *Mobile DNA*, 1st ed. Vol. 1 (Berg D. E. and Howe M. M., eds), pp. 593-617. Am. Soc. Microbiol., Washington DC, USA.
- Huxley C. and Gnirke A. (1991) Transfer of yeast artificial chromosomes from yeast to mammalian cells. *BioEssays* **13**, 545-550.
- Itzhaki J. E., Barnett M. A., MacCarthy A. B., Buckle V. J., Brown W. R. A., and Porter A. C. G. (1992) Targeted breakage of a human chromosome mediated by cloned human telomeric DNA. *Nat. Genet.* **2**, 283-287.
- Jagadeeswaran P., Tuan D., Forget B. G., and Weissman S. M. (1982) A gene deletion ending at the midpoint of a repetitive DNA sequence in one form of hereditary persistence of fetal haemoglobin. *Nature* **296**, 469-470.
- Jelinek W. R., Toomey T. P., Leinward L., Duncan C. H., Biro P. A., Choudary P. V., Weissman S. M., Rubin C. M., Houck C. M., Deininger P. L., and Schmid C. W. (1980) Ubiquitous, interspersed repeated sequences in mammalian genomes. *Proc. Natl. Acad. Sci. USA* **77**, 1398-1402.
- Jessberger R. and Berg P. (1991) Repair of deletions and double-strand gaps by homologous recombination in a mammalian *in vitro* system. *Mol. Cell. Biol.* **11**, 445-457.

- Jones D. S. C. and Schofield J. P. (1990) A rapid method for isolating high quality plasmid DNA suitable for DNA sequencing. *Nucleic Acids Res.* **18**, 7463-7464.
- Jurka J. (1993) A new family of recently retroposed human Alu repeats. *Nucleic Acids Res.* **21**, 2252.
- Jurka J. and Smith T. (1988) A fundamental division in the Alu family of repeated sequences. *Proc. Natl. Acad. Sci. USA* **85**, 4775-4778.
- Kaelbling M. and Klinger H. P. (1986) Suppression of tumorigenicity in somatic cell hybrids. *Cytogenet. Cell Genet.* **41**, 65-70.
- Kaina B., Fritz G., and Coquerelle T. (1991) Identification of human genes involved in repair and tolerance of DNA damage. *Radiation and Environmental Biophysics* **30**, 1-19.
- Kao F.-T. (1983) Somatic cell genetics and gene mapping, in *International Review of Cytology Vol. 85*, pp. 109-145. Academic Press Inc., .
- Kao F.-T. and Puck T. T. (1968) Genetics of somatic mammalian cells, VII. Induction and isolation of nutritional variants in Chinese hamster cells. *Proc. Natl. Acad. Sci. USA* **60**, 1275-1281.
- Kao F.-T., Jones C., and Puck T. T. (1976) Genetics of somatic mammalian cells: Genetic, immunologic, and biochemical analysis with Chinese hamster cell hybrids containing selected human chromosomes. *Proc. Natl. Acad. Sci. USA* **73**, 193-197.
- Kaufman R. J., Sharp P. A., and Latt S. A. (1983) Evolution of chromosomal regions containing transfected and amplified dihydrofolate reductase sequences. *Mol. Cell. Biol.* **3**, 699-711.
- Kazazian Jr H. H., Wong c, Youssoufian H., Scott A. F., Phillips D. G., and Antonarakis S. E. (1988) Haemophilia A resulting from de novo insertions of L1 sequences represents a novel mechanism for mutation in man. *Nature.* **332**, 164-166.
- Killary A. M., Wolf M. E., Giambernadi T. A., and Naylor S. L. (1992) Definition of a tumor suppressor locus within human chromosome 3p21-p22. *Proc. Natl. Acad. Sci. USA* **89**, 10877-10881.
- Klinger H. P. (1982) Suppression of tumorigenicity. *Cytogenet. Cell Genet.* **32**, 68-84.
- Klobutcher L. A. and Ruddle F. H. (1979) Phenotype stabilisation and integration of transferred material in chromosome-mediated gene transfer. *Nature.* **280**, 657-660.
- Klobutcher L. A. and Ruddle F. H. (1981) Chromosome mediated gene transfer, in *Annual Review of Biochemistry, 1st ed. Vol. 50* (Snell, Boyer, Meister, and Richardson, eds), pp. 533-554. Annual Reviews Inc., Palo Alto, C.A. (Review chapter on cmgt)
- Knudson A. G. (1971) Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. USA* **68**, 820-823.



- Koi M., Morita H., Shimizu M., and Oshimura M. (1989a) Construction of mouse A9 clones containing a single human chromosome (x/autosome translocation) via microcell fusion. *Jpn. J. Cancer Res.* **80**, 122-125.
- Koi M., Shimizu M., Morita H., Yamada H., and Oshimura M. (1989b) Construction of mouse A9 clones containing a single human chromosome tagged with neomycin-resistance gene via microcell fusion. *Jpn. J. Cancer Res.* **80**, 413-418.
- Koi M., Johnson L. A., Kalikin L. M., Little P. F. R., Nakamura Y., and Feinberg A. P. (1993) Tumor cell growth arrest caused by subchromosomal transferable DNA fragments from chromosome 11. *Science*. **260**, 361-364.
- Komatsu K., Kodama S., Okumura Y., Koi M., and Oshimura M. (1990) Restoration of radiation resistance in ataxia telangiectasia cells by the introduction of normal human chromosome 11. *Mutat. Res.* **235**, 59-63.
- Koob M. and Szybalski W. (1990) Cleaving Yeast and Escherichia coli Genomes at a Single Site. *Science*. **250**, 271-273.
- Koob M., Grimes E., and Szybalski W. (1988) Conferring Operator Specificity on Restriction Endonucleases. *Science*. **241**, 1084-1086.
- Korenberg J. R. and Rykowski M. C. (1988) Human genome organisation: Alu, Lines, and the molecular structure of metaphase chromosome bands. *Cell* **53**, 391-400.
- Kornreich R., Bishop D. F., and Desnick R. J. (1990) Alpha-galactosidase A gene rearrangements causing Fabry disease. *J. Biol. Chem.* **265**, 9319-9326.
- Krayev A. S., Kramerov D. A., Skryabin K. G., Ryskov A. P., Bayev A. A., and Georgiev G. P. (1980) The nucleotide sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse fold-back RNA. *Nucleic Acids Res.* **8**, 1201.
- Krowczynska A. M., Rudders R. A., and Krontiris T. G. (1990) The human minisatellite consensus at breakpoints of oncogene translocations. *Nucleic Acids Res.* **18**, 1121-1127.
- Kucherlapati R. S., Eves E. M., Song K.-Y., Morse B. S., and Smithies O. (1984) Homologous recombination between plasmids in mammalian cells can be enhanced by treatment of input DNA. *Proc. Natl. Acad. Sci. USA* **81**, 3153-3157.
- Kumar S. and Simons J. P. (1993) The effects of terminal heterologies on gene targeting by insertion vectors in embryonic stem cells. *Nucleic Acids Res.* **21**, 1541-1548.
- Kurdi-Haidar B., Levine F., Roemer K., LaPorte P., and Friedmann T. (1993) Provirus-anchored long-range (PAL) mapping of mammalian genomes. *Genomics*. **15**, 305-310.
- Laird P. W., Zijderveld A., Linders K., Rudnicki M. A., Jaenisch R., and Berns A. (1991) Simplified mammalian DNA isolation procedure. *Nucleic Acids Res.* **19**, 4293-4294.

- Lambert C., Schultz R. A., Smith M., Wagner-McPherson C., McDaniel L. D., Donlon T., Stanbridge E. J., and Friedberg E. C. (1991) Functional complementation of ataxia-telangiectasia group D (AT-D) cells by microcell-mediated chromosome transfer and mapping of the AT-D locus to the region 11q22-23. *Proc. Natl. Acad. Sci. USA* **88**, 5907-5911.
- Lawrence J. B., Villnave C. A., and Singer R. H. (1988) Sensitive, high-resolution chromatin and chromosome mapping *in situ*: presence and orientation of two closely integrated copies of EBV in a lymphoma line. *Cell* **52**, 51-61.
- Lee flank E. P., Liu W.-M., Hashimoto C., Choudray P. V., and Schmid C. W. (1992) Phylogenetic evidence for multiple Alu source genes. *J. Mol. Evol.* **35**, 7-16.
- Lehrman M. A., Russel D. W., Goldstein J. L., and Brown M. S. (1986) Exon-Alu recombination deletes 5 kilobases from the low density lipoprotein receptor gene, producing a null phenotype in familial hypercholesterolemia. *Proc. Natl. Acad. Sci. USA* **83**, 3679-3683.
- Lehrman M. A., Goldstein J. L., Russell D. W., and Brown M. S. (1987a) Duplication of seven exons in LDL receptor gene caused by Alu-Alu recombination in a subject with familial hypercholesterolemia. *Cell* **48**, 827-835.
- Lehrman M. A., Russell D. W., Goldstein J. L., and Brown M. S. (1987b) Alu-Alu recombination deletes splice acceptor sites and produces secreted low density lipoprotein receptor in a subject with familial hypercholesterolemia. *J. Biol. Chem.* **262**, 3354-3361.
- Leigh Brown A. (1984) On the origin of the Alu family of repeated sequences. *Nature*. **312**, 106.
- Levens D. and Howley P. M. (1985) Novel method for identifying sequence-specific DNA-binding proteins. *Mol. Cell. Biol.* **5**, 2307-2315.
- Lewis S., Gifford A., and Baltimore D. (1985) DNA elements are asymmetrically joined during site-specific recombination of kappa immunoglobulin chains. *Science*. **228**, 677-684.
- Li L. and Bray P. F. (1993) Homologous recombination among three intragene Alu sequences causes an inversion-deletion resulting in the hereditary bleeding disorder Glanzmann thrombasthenia. *Am. J. Hum. Genet.* **53**, 140-149.
- Lichter P. and Cremer T. (1992) Chromosome analysis by non-isotopic *in situ* hybridisation, in *Human Cytogenetics. A Practical Approach, 1st ed. Vol. 1, Constitutional Analysis* (Rooney D. E. and Czepulkowski B. H., eds), pp. 157-192. IRL Press, Oxford, UK.
- Liebold D. M., Swergold G. D., Singer M. F., Thayer R. E., Dombroski B. A., and Fanning T. F. (1990) Translation of LINE-1 DNA elements *in vitro* and in human cells. *Proc. Natl. Acad. Sci. USA* **87**, 6990-6994.
- Lin C. S., Goldthwait D. A., and Samols D. (1988) Identification of Alu transposition in human lung carcinoma cells. *Cell* **54**, 153-159.

- Lin F. L., Sperle K., and Sternberg N. (1985) Recombination in mouse L cells between DNA introduced into cells and homologous chromosomal sequences. *Proc. Natl. Acad. Sci. USA* **82**, 1391-1395.
- Lin F. L., Sperle K. M., and Sternberg N. L. (1987) Extrachromosomal recombination in mammalian cells as studied with single- and double-stranded DNA substrates. *Mol. Cell. Biol.* **7**, 129-140.
- Lin F. L., Sperle K., and Sternberg N. (1990a) Intermolecular recombination between DNAs introduced into mouse L cells is mediated by a nonconservative pathway that leads to crossover products. *Mol. Cell. Biol.* **10**, 103-112.
- Lin F. L., Sperle K., and Sternberg N. (1990b) Repair of double-stranded DNA breaks by homologous DNA fragments during transfer of DNA into mouse L cells. *Mol. Cell. Biol.* **10**, 113-119.
- Lin F.-L., Sperle K., and Sternberg N. (1984) Model for homologous recombination during transfer of DNA into mouse L cells: role for DNA ends in the recombination process. *Mol. Cell. Biol.* **4**, 1020-1034.
- Littlefield J. (1964) Selection of hybrids from mating of fibroblasts *in vitro* and their presumed recombinants. *Science*. **145**, 709-710.
- Lugo T. G., Handelin B., Killary A. M., Housman D. E., and Fournier R. E. K. (1987) Isolation of microcell hybrid clones containing retroviral vector insertions into specific human chromosomes. *Mol. Cell. Biol.* **7**, 2814-2820.
- Maio J. J., Brown F. L., McKenna W. G., and Muisch P. R. (1981) Toward a molecular paleontology of primate genomes, II: the KpnI families of alphoid DNAs. *Chromosoma (Berl)* **83**, 127-144.
- Maniatis T., Fritsch E. F., and Sambrook J. (1989) *Molecular Cloning: A laboratory manual*, 2nd ed. Cold Spring Harbor Publications, USA.
- Mansour S. L., Thomas K. R., and Capecchi M. R. (1988) Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature*. **336**, 348-352.
- Maraia R. J. (1991) The subset of mouse B1 (Alu-equivalent) sequences expressed as small processed cytoplasmic transcripts. *Nucleic Acids Res.* **19**, 5695-5702.
- Maraia R. J., Driscoll C. T., Bilyeu T., Hsu C., and Darlington G. J. (1993) Multiple dispersed loci produce small cytoplasmic Alu RNA. *Mol. Cell. Biol.* **13**, 4233-4241.
- Martin S. L. (1991) LINEs. *Curr. Op. Genet. Dev.* **1**, 505-508.
- Matera A. G., Hellmann U., Hintz M. F., and Schmid C. W. (1990a) Recently transposed Alu repeats result from multiple source genes. *Nucleic Acids Res.* **18**, 6019-6023.
- Matera A. G., Hellmann U., and Schmid C. W. (1990b) A transpositionally and transcriptionally competent Alu subfamily. *Mol. Cell. Biol.* **10**, 5424-5432.



- Mathias S. L., Scott A. F., Kazazian Jr H. H., Boeke J. D., and Gabriel A. (1991) Reverse transcriptase encoded by a human transposable element. *Science*. **254**, 1808-.
- Maule J. (1994) Electrophoretic karyotype analysis, in *Chromosome Analysis Protocols*, 1st ed. (Gosden J., ed) Humana Press Inc., New Jersey, USA.
- McBride O. W. and Ozer H. L. (1973) Transfer of genetic information by purified metaphase chromosomes. *Proc. Natl. Acad. Sci. USA* **70**, 1258-1262.
- McClelland M., Hanish J., Nelson M., and Patel Y. (1988) KGB: a single buffer for all restriction endonucleases. *Nucleic Acids Res.* **16**, 364.
- McClintock B. (1942) The fusion of broken ends of chromosomes following nuclear fusion. *Proc. Natl. Acad. Sci. USA* **28**, 458-463.
- Merry D. E., Pathak S., Hsu T. C., and Brinkley B. R. (1985) Anti-kinetochore antibodies: use as probes for inactive centromeres. *Am. J. Hum. Genet.* **37**, 425-430.
- Meselson M. S. and Radding C. M. (1975) A general model for genetic recombination. *Proc. Natl. Acad. Sci. USA* **72**, 358-361.
- Migeon B. R. and Miller C. S. (1968) Human-mouse somatic cell hybrids with single human chromosome (GroupE): link with thymidine kinase activity. *Science*. **162**, 1005-1006.
- Miki Y., Nishisho I., Horii A., Miyoshi Y., Utsunomiya J., Kinzler K. W., Vogelstein B., and Nakamura Y. (1992) Disruption of the APC gene by a retrotransposal insertion of L1 sequence in colon cancer. *Cancer. Res.* **52**, 643-645.
- Miles C. and Meuth M. (1989) Hamster Line and Alu-equivalent sequences are present in the small polydispersed circular DNA population of CHO cells. *Nucleic Acids Res.* **17**, 5846.
- Monteilhet C., Perrin A., Thierry A., Colleaux L., and Dujon B. (1990) Purification and characterisation of the *in vitro* activity of I-SceI, a novel and highly specific endonuclease encoded by a group I intron. *Nucleic Acids Res.* **18**, 1407-1412.
- Moyzis R. K., Albright K. L., Bartholdi M. F., Cram L. S., Deaven L. L., Hildebrand C. E., Joste N. E., Longmire J. L., Meyne J., and Schwarzbach-Robinson T. (1987) Human chromosome-specific repetitive DNA sequences: novel markers for genetic analysis. *Chromosoma* **95**, 375-386.
- Moyzis R. K., Torney D. C., Meyne J., Buckingham J. M., Wu J. R., Burks C., Sirotkin K. M., and Goad W. B. (1989) The distribution of interspersed repetitive DNA sequences in the human genome. *Genomics*. **4**, 273-289.
- Mulligan R. C. and Berg P. (1980) Expression of a bacterial gene in mammalian cells. *Science*. **209**, 1422.
- Mulligan R. C. and Berg P. (1981) Selection for animal cells that express the *Eschericia coli* gene coding for xanthine-guanine phosphoribosyltransferase. *Proc. Natl. Acad. Sci. USA* **76**, 2072-2076.

- Muratani K., Hada T., Yamamoto Y., Kaneko T., Shigeto Y., Ohue T., Furuyama J., and Higashino K. (1991) Inactivation of the cholinesterase gene by Alu insertion: Possible mechanism for human gene transposition. *Proc. Natl. Acad. Sci. USA* **88**, 11315-11319.
- Nabholz M., Miggiano V., and Bodmer W. (1969) Genetic analysis with human-mouse somatic cell hybrids. *Nature*. **223**, 358-363.
- Negrini M., Castagnoli A., Sabbioni S., Recanatini E., Giovannini G., Possati L., Stanbridge E. J., Nenci I., and Barbanti-Brodano G. (1992) Suppression of tumorigenesis by the breast cancer cell line MCF-7 following transfer of a normal human chromosome 11. *Oncogene* **7**, 2013-2018.
- Nickoloff J. A. (1992) Transcription enhances intrachromosomal homologous recombination in mammalian cells. *Mol. Cell. Biol.* **12**, 5311-5318.
- Noguchi T. (1991) Mini-maxi prep: an efficient and inexpensive procedure for medium scale preparation of high-quality supercoiled plasmid DNA. *Technique* **3**, 33-36.
- Okada Y. (1958) The fusion of Ehrlich's tumor cells caused by HJV virus in vitro. *Biken's J.* **1**, 103-110.
- Onno M., Nakamura T., Hillova J., and Hill M. (1992) Rearrangement of the human *trc* oncogene by homologous recombination between Alu repeats of nucleotide sequences from two different chromosomes. *Oncogene* **7**, 2519-2523.
- Oshimura M., Kugoh H., Koi M., Shimizu M., Yamada H., Satoh H., and Barrett J. C. (1990) Transfer of a normal human chromosome 11 suppresses tumorigenicity of some but not all tumor cell lines. *JCB* **42**, 135-142.
- Pachnis V., Pevny L., Rothstein R., and Constantini F. (1990) Transfer of a yeast artificial chromosome carrying human DNA from *Saccharomyces cerevisiae* into mammalian cells. *Proc. Natl. Acad. Sci. USA* **87**, 5109-5113.
- de Pagter-Holthuizen P., Jansen M., van Schaik F. M. A., van der Kammen R., Oosterwijk C., van der Brande J. L., and Sussenbach J. S. (1987) The human insulin-like growth factor 2 gene contains two development- specific promoters. *FEBS* **214**, 259-264.
- Paulson K. E. and Schmid C. W. (1986) Transcriptional inactivity of Alu repeats in HeLa cells. *Nucleic Acids Res.* **14**, 6145-6158.
- Pavan W. J. and Reeves R. H. (1991) Integrative selection of human chromosome-specific yeast artificial chromosomes. *Proc. Natl. Acad. Sci. USA* **88**, 7788-7791.
- Pavan W. J., Hieter P., and Reeves R. (1990) Generation of deletion derivatives by targeted transformation of human-derived yeast artificial chromosomes. *Proc. Natl. Acad. Sci. USA* **87**, 1300-1304.
- Pennington S. L. and Wilson J. H. (1991) Gene targeting in Chinese hamster ovary cells is conservative. *Proc. Natl. Acad. Sci. USA* **88**, 9498-9502.



- Perez-Stable C., Shen C. C., and Shen C.-K. J (1988) Enrichment and depletion of Hela topoisomerase I recognition sites among specific types of DNA elements. *Nucleic Acids Res.* **16**, 7975-7993.
- Persing D. H. and Landry M. L. (1989) *In vitro* amplification techniques for the detection of nucleic acids: new tools for the diagnostic laboratory. *The Yale Journal of Biology and Medicine* **62**, 159-171.
- Perucho M., Hanahan D., Lipsich L., and Wigler M. (1980) Isolation of chicken thymidine kinase gene by plasmid rescue. *Nature.* **285**, 207-210.
- Pinkel D., Straume T., and Gray J. W. (1986) Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridisation. *Proc. Natl. Acad. Sci. USA* **83**, 2934-2938.
- Ponder B. (1988) Gene losses in human tumours. *Nature.* **335**, 400-403.
- Pontecorvo G., Riddle P. N., and Hales A. (1977) Time and mode of fusion of human fibroblasts treated with polyethylene glycol (PEG). *Nature.* **265**, 257-258.
- Porteous D. J. (1987) Chromosome mediated gene transfer: a functional assay for complex loci and an aid to human genome mapping. *Trends. Genet.* **3**, 177-181.
- Porteous D. J., Bickmore W., Christie S., Boyd P. A., Cranston G., Fletcher J. M., Gosden J. R., Rout D., Seawright A., Simola K. O. J., van Heyningen V., and Hastie N. D. (1987) HRAS1-selected chromosome transfer generates markers that colocalize aniridia- and genitourinary dysplasia-associated translocation breakpoints and the Wilms tumor gene within band 11p13. *Proc. Natl. Acad. Sci. USA* **84**, 5355-5359.
- Pritchard C. and Goodfellow P. N. (1986) Development of new methods in human gene mapping: selection for fragments of the human Y chromosome after chromosome-mediated gene transfer. *Embo. J.* **5**, 979-985.
- Puck T. T. and Fao F. T. (1982) Somatic cell genetics and its application to medicine. *Ann.Rev.Genet.* **16**, 225-271.
- Puck T. T., Wuthier P., Jones C., and Kao F.-T. (1971) Genetics of somatic mammalian cells: lethal antigens as genetic markers for study of human linkage groups. *Proc. Natl. Acad. Sci. USA* **68**, 3102-3106.
- Pulciana S., Santos E., Lauver A. V., Long L. K., Robbins K. C., and Barbacid M. (1982) Oncogenes in human tumor cell lines: molecular cloning of a transforming gene from human bladder carcinoma cells. *Proc. Natl. Acad. Sci. USA* **79**, 2845-2849.
- Quentin Y. (1989) Successive waves of fixation of B1 variants in rodent lineage history. *J. Mol. Evol.* **28**, 299-305.
- Reeve A. E., Sih S. A., Raizes A. M., and Feinberg A. P. (1989) Loss of allelic heterozygosity at a second locus on chromosome 11 in sporadic wilm's tumor cells. *Mol. Biol. Evol.* **9**, 1799-1803.

- te Riele H., Maandag E. R., and Berns A. (1992) Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs. *Proc. Natl. Acad. Sci. USA* **89**, 5128-5132.
- Rogan P. K., Pan J., and Weissman S. M. (1987) L1 repeat elements in the human E-Y-globin gene intergenic region: sequence analysis and concerted evolution within this family. *Mol. Biol. Evol.* **4**, 327-342.
- Roth D. and Wilson J. (1988) Illegitimate recombination in mammalian cells, in *Genetic Recombination*, 1st ed. (Kucherlapati R. and Smith G. R., eds), pp. 621-652. American Society of Microbiology, Washington DC.
- Ruddle F. H. (1981) A new era in mammalian gene mapping: somatic cell genetics and recombinant DNA methodologies. *Nature*. **294**, 115-120.
- Sadler J. R., Sasmor H., and Betz J. L. (1983) A perfectly symmetric lac operator binds the lac repressor very tightly. *Proc. Natl. Acad. Sci. USA* **80**, 6785-6789.
- Sainz J., Pevny L., Wu Y., Cantor C. R., and Smith C. L. (1992) Distribution of interspersed repeats (Alu and Kpn) on NotI restriction fragments of human chromosome 21. *Proc. Natl. Acad. Sci. USA* **89**, 1080-1084.
- Saxon P. J. and Stanbridge E. J. (1987) Transfer and selective retention of single specific human chromosomes via microcell-mediated chromosome transfer, in *Methods in Enzymology Vol. 151*, pp. 313-325. Academic Press Inc., .
- Saxon P. J., Srivatsan E. S., Leipzig G. V., Sameshima J. H., and Stanbridge E. J. (1985) Selective transfer of individual human chromosomes to recipient cells. *Mol. Cell. Biol.* **5**, 140-146.
- Saxon P. J., Srivatsan E. S., and Stanbridge E. J. (1986) Introduction of human chromosome 11 via microcell transfer controls tumorigenic expression of HeLa cells. *Embo. J.* **5**, 3461-3466.
- Schedl A., Montoliu L., Kelsey G., and Schutz G. (1993) A yeast artificial chromosome covering the tyrosinase gene confers copy number-dependent expression in transgenic mice. *Nature*. **362**, 258-261.
- Schmid C. and Maraia R. (1992) Transcriptional regulation and transpositional selection of active SINE sequences. *Curr. Biol.* **2**, 874-882.
- Scott A. F., Schmeckpeper B. J., Abdelrazik M., Theisen Comey C., O'Hara B., Pratt Rossiter J., Cooley T., Heath P., Smith K. D., and Margolet L. (1987) Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. *Genomics*. **1**, 113-125.
- Shafit-Zagardo B., Brown F. L., Maio J. J., and Adams J. W. (1982) Kpn1 families of long, interspersed repetitive DNAs associated with the human  $\beta$ -globin gene cluster. *Gene* **20**, 397-407.
- Shafit-Zagardo B., Brown F. L., Zavodny P. J., and Maio J. J. (1983) Transcription of the Kpn1 families of long interspersed DNAs in human cells. *Nature*. **304**, 277-280.

- Shen M. R. and Deininger P. L. (1992) An *in vivo* assay for measuring the recombination potential between DNA sequences in mammalian cells. *Anal. Biochem.* **205**, 83-89.
- Shen M. R., Batzer M. A., and Deininger P. L. (1991) Evolution of the master Alu gene(s). *J. Mol. Evol.* **33**, 311-320.
- Shibasaki Y. (1993) The initial mechanism involved in N-myc oncogene amplification in human neuroblastoma cell lines. *Dr. Sc. Hum.* Heidelberg.
- Shih C. and Weinberg R. A. (1982) Isolation of a transforming sequence from a bladder carcinoma cell line. *Cell* **29**, 161-169.
- Shih C., Padhy L. C., Murray M., and Weinberg R. A. (1981) Transforming genes of carcinomas and neuroblastomas introduced into mouse fibroblasts. *Nature.* **290**, 261-264.
- Shimizu M., Yokota J., Mori N., Shuin T., Shinoda M., Terada M., and Oshimura M. (1990) Introduction of normal chromosome 3p modulates the tumorigenicity of a human renal cell carcinoma cell line YCR. *Oncogene* **5**, 185-194.
- Singer M. F. (1982) SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. *Cell* **28**, 433-434.
- Singer M. F., Thayer R. E., Grimaldi G., Lerman M. I., and Fanning T. G. (1983) Homology between the Kpn1 primate and BamH1 (M1F-1) rodent families of long interspersed repeated sequences. *Nucleic Acids Res.* **11**, 5739-5743.
- Singer R. H., Bentley Lawrence J., and Villnave C. (1986) Optimization of *in situ* hybridisation using isotopic and non-isotopic detection methods. *BioTech* **4**, 230.
- Skowronski J., Fanning T. G., and Singer M. F. (1988) Unit-length Line-1 transcripts in human teratocarcinoma cells. *Mol. Cell. Biol.* **8**, 1385-1397.
- Small J. and Scangos G. (1983) Recombination during gene transfer can restore the function of deleted genes. *Science.* **219**, 174-176.
- Smith G. R. (1983) Chi hotspots of generalised recombination. *Cell* **34**, 709-710.
- Smith G. R. and Stahl F. W. (1985) Homologous recombination promoted by Chi sites and RecBC enzyme of *Escherichia coli*. *BioEssays* **2**, 244-249.
- Smith K. A., Gorman P. A., Stark M. B., Groves R. P., and Stark G. R. (1990) Distinctive chromosomal structures are formed very early in the amplification of CAD genes. *Cell* **63**, 1219-1227.
- Smith K. A., Stark M. B., Gorman P. A., and Stark G. R. (1992) Fusions near telomeres occur very early in the amplification of CAD genes in Syrian hamster cells. *Proc. Natl. Acad. Sci. USA* **89**, 5427-5431.
- Smithies O., Gregg R. G., Boggs S. S., Koralewski M. A., and Kucherlapati R. S. (1985) Insertion of DNA sequences into the human chromosomal  $\beta$ -globin locus by homologous recombination. *Nature.* **317**, 230-234.



- Song K.-Y., Chekuri L., Rauth S., Ehrlich S., and Kucherlapati R. (1985) Effect of double-strand breaks on homologous recombination in mammalian cells and extracts. *Mol. Cell. Biol.* **5**, 3331-3336.
- Soriano P., Meunier-Rotival M., and Bernardi G. (1983) The distribution of interspersed repeats is nonuniform and conserved in the mouse and human genomes. *Proc. Natl. Acad. Sci. USA* **80**, 1816-1820.
- Southern E. M. (1975) Detection of specific sequences amongst DNA fragments separated by agarose gel electrophoresis. *J. Mol. Biol.* **98**, 503-517.
- Southern P. J. and Berg P. (1982) Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early promoter region. *J. Virol.* **1**, 327-341.
- Srivatsan E. S., Stanbridge E. J., Saxon P. J., Stambrook P. J., Trill J. J., and Tischfield J. A. (1984) Plasmid, phage and genomic DNA-mediated transfer and expression of prokaryotic and eukaryotic genes in cultured human cells. *Cytogenet. Cell Genet.* **38**, 227-234.
- Stanbridge E. J. (1976) Suppression of malignancy in human cells. *Nature*. **260**, 17-20.
- Stanbridge E. J. (1992) Functional evidence for human tumor suppressor genes: chromosome and molecular genetic studies, in *Tumor Suppressor Genes, the Cell Cycle and cancer, Cancer Surveys ed. Vol. 12*, pp. 5-24. Imperial Cancer Research Fund, .
- Stoppa-Lyonnet D., Carter P. E., Meo T., and Tosi M. (1990) Clusters of intragenic Alu repeats predispose the human C1 inhibitor locus to deleterious rearrangements. *Proc. Natl. Acad. Sci. USA* **87**, 1551-1555.
- Strathdee C. A., Gavish H., Shannon W. R., and Buchwald M. (1992) Cloning of cDNAs for Fanconi's anaemia by functional complementation. *Nature*. **356**, 763-767.
- Strauss W. M. and Jaenisch R. (1992) Molecular complementation of a collagen mutation in mammalian cells using yeast artificial chromosomes. *Embo. J.* **11**, 417-422.
- Sumner A. T., Evans H. J., and Buckland R. A. (1971) New technique for distinguishing between human chromosomes. *Nature*. **232**, 31-32.
- Szostak J. W., Orr-Weaver T. L., Rothstein R. J., and Stahl F. W. (1983) The double-strand-break repair model for recombination. *Cell* **33**, 25-35.
- Tanaka K., Satokata I., Ogita Z., Uchida T., and Okada Y. (1989) Molecular cloning of a mouse DNA repair gene that complements the defect of group-A xeroderma pigmentosum. *Proc. Natl. Acad. Sci. USA* **86**, 5512-5516.
- Tanaka K., Oshimura M., Kikuchi R., Seki M., Hayashi T., and Miyaki M. (1991) Suppression of tumorigenicity in human colon carcinoma cells by introduction of normal chromosome 5 or 18. *Nature*. **349**, 340-342.

- Tanaka K., Yanoshita R., Konishi M., Oshimura M., Maeda Y., Mori T., and Miyaki M. (1993) Suppression of tumorigenicity in human colon carcinoma cells by introduction of normal chromosome 1p36 region. *Oncogene* **8**, 2253-2258.
- Thierry A. and Dujon B. (1992) Nested chromosomal fragmentation in yeast using the meganuclease I-SceI: a new method for physical mapping of eukaryotic chromosomes. *Nucleic Acids Res.* **20**, 5625-5631.
- Thomas B. J. and Rothstein R. (1989) Elevated recombination rates in transcriptionally active DNA. *Cell* **56**, 619-630.
- Thomas K. R. and Capecchi M. R. (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**, 503-512.
- Thomas K. R., Folger K. R., and Capecchi M. R. (1986) High frequency targeting of genes to specific sites in the mammalian genome. *Cell* **44**, 419-428.
- Thomas K. R., Deng C., and Capecchi M. R. (1992) High-fidelity gene targeting in embryonic stem cells by using sequence replacement vectors. *Mol. Cell. Biol.* **12**, 2919-2923.
- Thompson S., Clarke A. R., Pow A. M., Hooper M. L., and Melton D. W. (1989) Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. *Cell* **56**, 313-321.
- Thweatt R. and Goldstein S. (1993) Werner syndrome and biological ageing: a molecular genetic hypothesis. *BioEssays* **15**, 421-425.
- Trask B. J. (1991) DNA sequence localization in metaphase and interphase cells by fluorescence *in situ* hybridization, in *Methods in Cell Biology*, 1st ed. Vol. 35, pp. 3-35. Academic Press Inc., .
- Trent J. M., Stanbridge E. J., McBride H. L., Meese E. U., Casey G., Araujo D. E., Witkowski C. M., and Nagle R. B. (1990) Tumorigenicity in human melanoma cell lines controlled by introduction of human chromosome 6. *Science*. **247**, 568-571.
- Tunnacliffe A., Jones C., and Goodfellow P. (1983) Somatic cell genetics, immunogenetics and gene mapping. *Immunol. Today* **4**, 230-233.
- Ullu E. and Tschudi C. (1984) Alu sequences are processed 7SL RNA genes. *Nature*. **312**, 171-172.
- Valancius V. and Smithies O. (1991) Double-strand gap repair in a mammalian gene targeting reaction. *Mol. Cell. Biol.* **11**, 4389-4397.
- Vieira J. and Messing J. (1982) The pUC plasmids, and M13mp7-derived system for insertion mutagenesis and sequencing with universal primers. *Gene* **19**, 259-269.
- Voliva C. F., Martin S. L., Hutchinson III C. A., and Edgell M. H. (1984) Dispersal process associated with the L1 family of interspersed repetitive DNA sequences. *J. Mol. Biol.* **178**, 795-813.



- Wagner M. J., Sharp J. A., and Summers W. C. (1981) Nucleotide sequence of the thymidine kinase gene of herpes simplex virus type 1. *Proc. Natl. Acad. Sci. USA* **78**, 1441-1445.
- Wake C. T., Vernaleone F., and Wilson J. H. (1985) Topological requirements for homologous recombination among DNA molecules transfected into mammalian cells. *Mol. Cell. Biol.* **5**, 2080-2089.
- Waldman A. S. and Liskay R. M. (1988) Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. *Mol. Cell. Biol.* **8**, 5350.
- Waldren C., Braaton M., Vannais D., Fouladi B., and Parker R. D. (1992) The use of human repetitive DNA to target selectable markers into the human chromosome of a human-hamster hybrid cell line (AL). *Somat. Cell Mol. Genet.* **18**, 417-422.
- Wallace M. R., Andersen L. B., Saulino A. M., Gregory P. E., Glover T. W., and Collins F. S. (1991) A *de novo* Alu insertion results in neurofibromatosis type 1. *Nature*. **353**, 864-866.
- Wallenburg J. C., Nepveu A., and Chartrand P. (1987) Integration of a vector containing rodent repetitive elements in the rat genome. *Nucleic Acids Res.* **15**, 7849-7862.
- Walter M. A. and Goodfellow P. N. (1993) Radiation hybrids : irradiation and fusion gene transfer. *Trends. Genet.* **9**, 352-356.
- Walter P. and Blobel G. (1982) Signal recognition particle contains a 7S RNA essential for protein translocation across endoplasmic reticulum. *Nature*. **299**, 691-698.
- Weber C. A., Salazar E. P., Stewart S. A., and Thompson L. H. (1988) Molecular cloning and biological characterization of a human gene, ERCC2, that corrects the nucleotide excision-repair defect in CHO UV5 cells. *Mol. Cell. Biol.* **8**, 1137-1146.
- Weeda G., van Ham R. C. A., Vermeulen W., Bootsma D., van der Eb A. J., and Hoeijmakers J. H. J. (1990) A presumed helicase encoded by ERCC-3 is involved in human repair disorders Xeroderma pigmentosum and Cockayne's syndrome. *Cell* **62**, 777-791.
- Weeda G., Hoeijmakers J. H. J., and Bootsma D. (1993) Genes controlling nucleotide excision repair in eukaryotic cells. *BioEssays* **15**, 249-258.
- Weiss M. C. and Green H. (1967) Human-mouse hybrid cell lines containing partial components of human chromosomes and functioning human genes. *Proc. Natl. Acad. Sci. USA* **58**, 1104-1111.
- Weissenbach J., Gyapay G., Dib C., Vignal A., Morissette J., Millasseau P., Vaysseix G., and Lathrop M. (1992) A second-generation linkage map of the human genome. *Nature*. **359**, 794-801.
- Weissman B. E., Saxon P. J., Pasquale S. R., Jones G. R., Geiser A. G., and Stanbridge E. J. (1987) Introduction of a normal human chromosome 11 into a Wilms' tumor cell line controls its tumorigenic expression. *Science*. **236**, 175-180.

- Westerveld A., Hoeijmakers J. H. J., van Duin M., de Wit J., Odijk H., Pastink A., Wood R. D., and Bootsma D. (1984) Molecular cloning of a human DNA repair gene. *Nature*. **310**, 425-429.
- Wigler M., Silverstein S., Lee Lih-S., Pellicer A., Cheng Y.-C., and Axel R. (1977) Transfer of purified herpes simplex virus thymidine kinase gene to cultured mouse cells. *Cell* **11**, 223-232.
- Wigler M., Sweet R., Sim G. K., Wold B., Pellicer A., Lacy E., Maniatis T., Silverstein S., and Axel R. (1979) Transformation of mammalian cells with genes from prokaryotes and eukaryotes. *Cell* **16**, 777-785.
- Willard H. F. and Wayne J. S. (1987) Hierarchical order in chromosome-specific human alpha satellite DNA. *Trends. Genet.* **3**, 192-198.
- Windle B., Draper B. W., Yin Y., O'Gorman S., and Wahl G. M. (1991) A central role for chromosome breakage in gene amplification, deletion formation, and amplicon integration. *Genes and Dev.* **5**, 160-174.
- Windle B. E. and Wahl G. M. (1992) Molecular dissection of mammalian gene amplification: new mechanistic insights revealed by analyses of very early events. *Mutation Research* **276**, 199-224.
- Woodcock D. M., Crowther P. J., Doherty J., Jefferson S., DeCruz E., Noyer-Weidner M., Smith S. S., Micheal M. Z., and Graham M. W. (1989) Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. *Nucleic Acids Res.* **17**, 3469-3478.
- Xu W., Gorman P. A., Rider S. H., Hedge P. J., Moore G., Prichard C., Sheer D., and Solomon E. (1988) Construction of a genetic map of human chromosome 17 by use of chromosome-mediated gene transfer. *Proc. Natl. Acad. Sci. USA* **85**, 8563-8567.
- Yurov Y. B., Mitkevich S. P., and Alexandrov I. A. (1987) Application of cloned satellite DNA sequences to molecular-cytogenetic analysis of constitutive heterochromatin heteromorphisms in man. *Hum. Genet.* **76**, 157-164.
- Zeng H. and Wilson J. H. (1990) Gene targeting in normal and amplified cell lines. *Nature*. **344**, 170-173.